

N83-26155- N83-26175

NASA Conference Publication 2261

# National Conference on Energy Resource Management

## Proceedings

### Vol. I – Techniques, Procedures & Data Bases

*Proceedings of a working group meeting  
held at Baltimore, Maryland  
September 9-12, 1982*

**NASA**



**NASA FILE COPY**

*NASA Conference Publication 2261*

# National Conference on Energy Resource Management

## Vol. I – Techniques, Procedures & Data Bases

Theme: Integration of Remotely Sensed Data with Geographic Information  
Systems for Application in Energy Resource Management

Proceedings of a working group meeting  
Sponsored by Energy Planning Division  
A Division of the American Planning Association  
National Aeronautics and Space Administration  
Nuclear Regulatory Commission  
and U.S. Region of the Remote Sensing Society  
September 9-12, 1982



**NASA**  
National Aeronautics  
and Space Administration  
Scientific and Technical  
Information Branch



**Page intentionally left blank**

**Page intentionally left blank**



---

ENERGY RESOURCE MANAGEMENT

---

PROCEEDINGS OF THE NATIONAL CONFERENCE ON ENERGY RESOURCE MANAGEMENT  
Integration of Remotely Sensed Data with Geographic Information  
Systems for Application in Energy Resource Management

VOLUME I:        TECHNIQUES, PROCEDURES AND DATA BASES

VOLUME II:      APPLICATIONS

Edited by James O. Brumfield and Yale M. Schiffman

All papers submitted for the presentation at the National Conference on Energy Resource Management for the subsequent publication in these Proceedings have been reviewed, first on the basis of the abstracts and then on the basis of the final manuscripts. The authors copies were submitted camera ready and are presented essentially as we received them. Several full length papers were not received in time for inclusion thus we have taken the liberty of including the abstract. You may wish to contact the authors of those abstracts directly for further information. Inclusion of the paper in the Proceedings in no way constitutes an endorsement by the American Planning Association's Energy Planning Division nor by the sponsors of this conference of the authors' views and opinions.



**Page intentionally left blank**

**Page intentionally left blank**



---

#### CONFERENCE COMMITTEE

---

Yale M. Schiffman, Schiffman Energy Services, Inc.  
Conference Chairperson

Professor James O. Brumfield, Marshall University,  
Conference Co-chairperson; Regional Meetings Secretary,  
Remote Sensing Society

William J. Campbell, National Aeronautics and Space  
Administration, ERRSAC, Technical Project Officer/  
Conference Co-chairperson

Monique Gormont, APA Energy Planning Division,  
Conference Coordinator

Germain LaRouche, Nuclear Regulatory Commission

Dr. Dana Hall, The MITRE Corporation

Robert Tanenhaus, Implementation Management International

Marty McClure, AEROS Data Corporation

Vince Robinson, City University of New York; Regional  
Meetings Secretary, Remote Sensing Society

Pamela Moore, APA Energy Planning Division,  
Assistant Conference Coordinator

---

#### ABSTRACT AND PAPER REVIEW COMMITTEE

---

Dr. Dana Hall, The MITRE Corporation

Vince Robinson, City University of New York; Regional  
Meetings Secretary, Remote Sensing Society

Dr. Germain LaRouche, Nuclear Regulatory Commission

Robert Tanenhaus, Implementation Management International

Professor James O. Brumfield, Marshall University,  
Conference Co-chairperson; Regional Meetings Secretary,  
Remote Sensing Society

William J. Campbell, National Aeronautics and Space  
Administration, ERRSAC, Technical Project Officer/  
Conference Co-chairperson

---

TENTATIVE SCHEDULE OF FUTURE MEETINGS

---

August 23 - 27, 1983	Hyatt Regency, San Francisco, CA
January/February, 1984	Rio de Janiero, Brazil
September, 1984	Boston, MA
April/May, 1985	London, England (in conjunction with International Remote Sen- sing Society Annual Meeting)
September, 1985	San Francisco, CA

---

## Pre-Conference Workshops

---

Four pre-conference workshops were offered on the afternoon of September 9, 1982. These workshops were designed to broaden the understanding of participants in the areas of remote sensing, geographic information systems, energy resource management and energy facility siting. They lasted approximately 2 hours.

### I. ENERGY RESOURCE MANAGEMENT

Yale M. Schiffman, Schiffman Energy Services

This workshop provided a broad overview of energy resource management and explored the range of data needs for implementation of this type of a program in residential/commercial, industrial and utility sectors.

### II. REMOTE SENSING

Dr. Nicholas M. Short, NASA/Goddard Space Flight Center

This workshop provided an introduction and overview of remote sensing; its history, and what it is today. The electromagnetic spectrum, spectral signatures, and basic energy relationships were discussed. Photographic and non-photographic remote sensing systems were also explored along with several image analysis options.

### III. GEOGRAPHIC INFORMATION SYSTEMS

Chuck Killpack, IRIS International, Inc.

William J. Campbell, NASA/Goddard Space Flight Center

This workshop explored the concepts underlying Geographic Information Systems. GIS data bases were discussed. Emphasis was placed on guidelines for establishing a practical GIS system for energy resource management.

### IV. ENERGY FACILITY SITING

Rod Heller, Wirth Associates

Current practices by the industry on approaches to site screening, evaluation and selection were discussed. The types of data needed, traditional sources and limitations of different systems were also explored.



**Page intentionally left blank**

**Page intentionally left blank**

## TABLE OF CONTENTS

Preface	xvii
Acknowledgements	xxiii

### VOLUME I - REMOTE SENSING TECHNIQUES AND PROCEDURES

<u>Part 1</u> : Use of Remote Sensing and Geographic Information Systems for Integration and Modeling	1
A Study of Feature Extraction Using Divergence Analysis of Texture Features	3
W. Hallada, B. Bly and R. Boyd, Computer Sciences Corporation; S. Cox, NASA/Goddard Space Flight Center	
Principal Components as a Data Reduction and Noise Reduction Technique	19
M. L. Imhoff, W. J. Campbell, NASA/Goddard Space Flight Center	
Optimization of a Non-Traditional Unsupervised Classification Approach for Land Cover Analysis	32
R. K. Boyd, Computer Sciences Corp., NASA/Goddard Space Flight Center; J. O. Brumfield, Marshall University; W. J. Campbell, NASA/Goddard Space Flight Center	
Using HCM Thermal Data to Improve Classification of MSS Data	45
R. G. Witt, NASA/Goddard Space Flight Center; R. S. Sekhon, Computer Sciences Corporation/Goddard Space Flight Center	
The Use of Principal Components for Creating Improved Imagery for Geometric Control Point Selection	56
M. L. Imhoff, NASA/Goddard Space Flight Center	
Ground Truth Sampling and Landsat Accuracy Assessment	69
J. Robinson, Computer Sciences Corporation/Goddard Space Flight Center	
Landsat Digital Data Base Preparation for the Pennsylvania Defoliation Application Pilot Test	83
R. G. McLeod, A. Zobrist, California Institute of Technology, Jet Propulsion Laboratory	

Image Analysis for Facility Siting: A Comparison of Low- and High-Altitude Image Interpretability for Land Use/Land Cover Mapping	97
H. M. Borella, EG&G, Santa Barbara Operations; J. E. Estes, C. E. Ezra, J. Scepan, and L. R. Tinney, University of California, Santa Barbara	
Potential of a New Technique for Remote Sensing of Hydrocarbon Accumulations and Blind Uranium Deposits: Buried LiF Thermoluminescence Dosimeters	122
F. R. Siegel, Department of Geology, The George Washington University; J. E. Vaz, Instituto Venezolano de Investigaciones Cientificas, Caracas, Venezuela; R. C. Lindholm, Department of Geology, The George Washington University	
<u>Part 2:</u> Integration of Remotely Sensed and Other Geo-Referenced Data Bases into GIS for Modeling and Applications	133
GIS Integration for Quantitatively Determining the Capabilities of Five Remote Sensors for Resource Exploration	135
R. Pascucci and A. Smith, Autometric, Inc.	
ORSER Landsat Data Base of Pennsylvania	149
B. J. Turner, Office for Remote Sensing of Earth Resources, The Pennsylvania State University; D. L. Williams, NASA/Goddard Space Flight Center	
Merging Landsat Derived Land Covers into Quad-Referenced Geographic Information Systems	161
J. M. White, R. M. Ragan, University of Maryland, Remote Sensing Systems Laboratory; K. P. Lade, Department of Sociology and Anthropology, Salisbury State College	
Remote Sensing/GIS Integration for Site Planning and Resource Management	182
J. D. Fellows, University of Maryland, Department of Civil Engineering	
An Interface for Remote Sensing Digital Image Systems and Geographic Information Systems	203
R. R. Irish and W. L. Myers, The Pennsylvania State University	
Development of Landsat Derived Forest Cover Information for Integration into Adirondack Park GIS	212
R. Curran and J. Banta, Adirondack Park Agency	

Interactive Management and Updating of Spatial Data Bases	231
P. French and M. Taylor, Resources Planning Associates, Inc.	
The Oklahoma Geographic Information Retrieval System	239
W. A. Blanchard, Center for Applications of Remote Sensing, Oklahoma State University	
Data Base Management for Geographic Information Systems	250
M. G. Pavlides, Greenhorne & O'Mara, Inc.	
Spatially Characterizing Effective Timber Supply	256
J. Berry, Yale University, School of Forestry and Environmental Studies; J. Sailor, Inter-graph Corporation	
Cartographic Modeling: Computer-Assisted Analysis of Spatially Defined Neighborhoods	267
J. K. Berry, Yale University, School of Forestry and Environmental Studies; C. D. Tomlin, Harvard University, Graduate School of Design	
<u>Part 3:</u> Techniques, Procedures and Data Bases: Poster Sessions	279
Preprocessing of Thematic Mapper Simulator (TMS) Image Data	281
F. J. Gunther, Computer Sciences Corporation; B. G. Bly, Computer Sciences Corporation; and W. J. Campbell, NASA/Goddard Space Flight Center	
Determination of Remote Sensor Capability by Means of an Automated Geographic Information System	283
R. Pascucci and A. Smith, Autometric, Inc.	
Applicability, Cost, and Accuracy Comparisons of Several Change-Detection Digital Procedures	285
F. J. Gunther, Computer Sciences Corporation and W. J. Campbell, NASA/Goddard Space Flight Center	

## VOLUME II - APPLICATIONS

<u>Part 1:</u> Use of Remotely Sensed Data & GIS for Energy and Environmental Resource Management	287
Comparison of Land Cover Information from Landsat MSS and Airborne TMS for Hydrological Applications: Preliminary Results J. C. Gervin, NASA/Goddard Space Flight Center; Y. C. Lu, W. A. Hallada, and R. F. Marcell, Computer Sciences Corporation/Goddard Space Flight Center	289
The Role of Remote Sensing in Assessing Forest Biomass in Appalachian South Carolina W. Shain and L. Nix, Department of Forestry, Clemson University	303
An Integration of a GIS with Peatland Management J. C. Hoshal, Minnesota Department of Energy, Planning and Development; R. L. Johnson, Minnesota Department of Natural Resources, Division of Minerals	319
Geographic Information Systems for Assessing Existing and Potential Bio-Energy Resources: Their Use in Determining Land Use and Management Options which Minimize Ecological and Landscape Impacts in Rural Areas A. E. Jackman, J. G. Fabos and C. C. Carlozzi, Department of Landscape Architecture and Regional Planning and Department of Forestry, Fishery and Wildlife, University of Massachusetts	328
Assessment of Coastal Deterioration Using Historical Photography and a GIS M. Dosier and C. Sasser, Coastal Ecology Laboratory; J. Hill, Department of Civil Engineering, Louisiana State University	351
Landsat Applications for the State of Delaware J. C. Gervin, NASA; W. A. Hallada, Computer Sciences Corporation; R. G. Witt, NASA/Goddard Space Flight Center	353
<u>Part 2:</u> Energy Facility Siting	371
Impact of Demographic Siting Criteria and Environmental Suitability on Land Availability for Nuclear Reactor Siting K. L. Hansen, Dames & Moore	373



Multidimensional Programming Methods for Energy Facility Siting: Alternative Approaches B. D. Solomon, West Virginia University; K. E. Haynes, Indiana University	393
Airphoto Interpretation and the Selection of a Powerline Right-of-Way in Vermont H. Klunder, Hans Klunder Associates; R. B. Arend, Photographic Interpretation Corporation	410
Potential Role of Land Use and Land Cover Information in Powerplant Siting: Example of Three Mile Island J. R. Wray, U. S. Geological Survey	427
Computer-Aided Siting of Coal Conversion Facilities: A Case Study D. D. Moreno, Dames & Moore	432
<b><u>Part 3:</u> Uses of Remote Sensing for Reclamation and Surface Mining</b>	<b>447</b>
Mining Operations and Remote Sensing A. Anderson, Office of Surface Mining, U. S. Department of the Interior	449
Remote Sensing Applications to the Pennsylvania Abandoned Mine Reclamation Program E. Clemens and L. Warnick, Earth Satellite Corporation	450
A Temporal Approach to Monitor Surface Mine Reclamation Progress Via Landsat A. L. Davis, Ohio University; H. L. Bloemer, Ohio University; and J. O. Brumfield, Marshall University	457
Ohio's Abandoned Mine Lands Reclamation Program: A Study of Data Collection and Evaluation Techniques S. L. Sperry, NUS Corporation	463
<b><u>Part 4:</u> Remote Sensing and Geographic Information Systems for Resource Management: An International Perspective</b>	<b>479</b>
Remote Sensing--An International Perspective M. McClure, AEROS Data Corporation	487
Geographic Information System in Bolivia: A Case Study for Latin America P. M. Adrien, Inter-American Development Bank	489

Factors in the Effective Utilization of a Landsat Related Inventory in West Africa L. Hall, Earth Satellite Corporation	500
Potential Applications of Landsat Data in Energy Resource Management Associated with Kenya's Forests W. M. Maghenda, H. L. Bloemer, Ohio University; J. O. Brumfield, Marshall University	508
A Method for Peat Inventory Based on Landsat Data and Computerized Mapping S. Pala, Ontario Centre for Remote Sensing, Ministry of Natural Resources	518
<b>Part 5: Applications Forum &amp; Symposia</b>	535
Remote Sensing and Geographic Information Systems for Environmental Information Needs W. J. Campbell, NASA/Goddard Space Flight Center; R. Ballard, Nuclear Regulatory Commission	537
Data Integration/Remote Sensing at Los Alamos National Laboratory: An Update S. B. Freeman, S. L. Bolivar, and T. A. Weaver, Los Alamos National Laboratory	541
Energy Remote Sensing Applications Projects at the NASA-Ames Research Center S. D. Norman, W. C. Likens, and D. A. Mouat, NASA-Ames Research Center	545
A Methodology for Assessing the Regional Transportation Energy Demands of Different Spatial Residential Develop- ment Scenarios: A Case Study for the Upper Housatonic River Basin, Massachusetts J. A. Oski, J. Gy. Fabos, and M. Gross, Department of Landscape Architecture and Regional Planning, University of Massachusetts	552
Assessment Planning and Evaluation of Renewable Energy Resources: An Interactive Computer Assisted Procedure T.W. Aston, J. Gy. Fabos and E.B. MacDougall Department of Landscape Architecture, University of Massachusetts	566
Seattle's System for Evaluating Energy Options P. Logie and M. J. Macdonald, City Light Depart- ment, City of Seattle, Washington	589

Use of Remote Sensing for Environmental Analysis C. T. Cushwa, Fish and Wildlife Service; G. LaRouche, Nuclear Regulatory Commission; C. DuBrock, Biometrician, Pennsylvania Game Commission	597
Air Force Energy Requirements B. R. Lenz, HQ USAF/LEYSF, Washington, D.C.	615
Landsat Users Forum R. Koffler and D. Cotter, Department of Com- merce, National Oceanic and Atmospheric Administration	616
Low Cost Laser Disk for Geo-Referenced Data Storage W. J. Campbell, NASA/Goddard Space Flight Center	617
<u>Part 6: Applications: Poster Sessions</u>	619
Solar Technologies: When is the Resource Cost Too High Y. M. Schiffman, Schiffman Energy Services, Inc.	621
Practical Applications Using a High Resolution Infrared Imaging System D. W. Baraniak, Donohue and Associates	622
Integration of Fish and Wildlife Data with Remotely Sensed Land Use/Land Cover Data--and Demonstration Using Sites in Pennsylvania C. T. Cushwa, U. S. Fish and Wildlife Service; G. LaRouche, U. S. Nuclear Regulatory Commission; and C. DuBrock, Pennsylvania Game Commission	628
The Nuclear Regulatory Commission Emergency Planning Mapping Program A. Sinisgalli, E. Weinstein, and J. Coulson, U. S. Nuclear Regulatory Commission	629
Glossary of Terms	631
List of Exhibitors and Publishers	633
Biographical Sketches of Contributors	637
Conference Participants	661

**Page intentionally left blank**

**Page intentionally left blank**

## P R E F A C E

These proceedings contain a selection of papers presented at the National Conference on Energy Resource Management which was held at the Baltimore Hilton Hotel, Baltimore, Maryland, September 9-12, 1982. The papers cover a wide variety of subject areas related to the conference theme, "Integration of Remotely Sensed Data With Geographic Information Systems for Application in Energy Resource Management," and describe the current trends and advances in the application of these systems to a number of energy concerns.

The APA Energy Planning Division co-sponsored the National Conference on Energy Resource Management with the National Aeronautics and Space Administration, the Nuclear Regulatory Commission, and the U.S. Region of the Remote Sensing Society. The conference brought together for the first time a number of professionals in such diverse fields as remote sensing, geographic information systems, information systems, urban and regional planning, fish and wildlife management, geography, cartography, systems analysis and resource extraction to name a few. The audience also had an international flavor with representatives from India, South America, Africa, Europe, Canada, Asia, and the United States. In all, nearly two hundred professionals met to exchange information and ideas regarding the information needs to manage energy and natural resources that are used as feedstock, materials or special resources in support of energy development. Seventeen exhibitors displayed some of the latest hardware, software and services for use in resource management.

On September 9, several preconference tutorials were given to provide a more detailed understanding of remote sensing, geographic information systems, energy resource management, and facility siting. September 10 was the opening

session of the conference which was chaired by Yale M. Schiffman, President of the Energy Planning Division. Sharing the responsibilities for running the meeting was Dr. James Brumfield, Director of the Marshall University Remote Sensing Program and Mr. William J. Campbell, Project Manager ERRSAC, NASA Goddard, Space Flight Center. Dr. John MacElroy, Assistant Administrator for Satellites, NOAA, U.S. Department of Commerce was the keynote speaker. Dr. MacElroy took this opportunity to make the first public announcement of NOAA's plans to commercialize the Landsat Satellite Program and also mentioned that similar plans are being evaluated for the nation's weather satellites. The closing session of the conference held Sunday, September 12, was chaired by Dr. Phil Cressy, Head of ERRSAC/NASA, Dr. Richard Kott, Senior Staff Member--U.S. DOE, Dr. Ray Harris, Honorable General Secretary of the Remote Sensing Society, Mr. Schiffman and Dr. Brumfield.

The meeting closed on a positive note indicating that that there was indeed a need to continue dialog by specialists from this wide array of interests. The emphasis should be on improving the communication between the user community and the software, hardware and analytical support specialists in the field. Thus, this meeting serves as the starting point for a continuing series on the subject, the next which will be held in San Francisco, August 23-27, at the Hyatt Regency at Embarcadero and in early 1984 in Rio de Janeiro, Brazil.

The interdisciplinary oriented conference provided a forum for presenting and discussing scientific works in the areas of energy resource management, remote sensing, geographic information systems, other georeferenced data systems, environmental analysis and applied systems research. Nearly 200



scientists, engineers, planners, and other professionals from eight countries contributed to these proceedings which were held over a three-day period, and in which nearly 100 presentations were given, most of which are included in these proceedings. These papers will undergo an additional rigorous review by our editorial committee. The purpose of the additional review is to select a limited number of papers for inclusion in a state of the art, hard covered publication that will be published in 1983.

Since the main theme of the conference was "The Integration of Remotely Sensed Data With Geographic Information Systems for Application in Energy Resource Management," a large number of papers are included in Volume I and II that deal with this topic. The proceedings have been organized along subject lines rather than in order of presentation. The editors felt the information would be more useful to the readers if organized in this fashion. These papers examine, in Volume I, the techniques and procedures that have been used to integrate remotely sensed data with geographic information systems, while in Volume II the papers explore the topic from an applications focus. Many papers also explore the integration of remotely sensed data with other georeferenced data systems. The proceedings clearly reflect the trends to integrate remotely sensed data with a number of different georeferenced data systems and illustrate an emerging interest among a number of different specialties in energy resource management for use of these integrated data systems in environmental assessment, facility siting, and facility planning.

There also appears to be a strong interest in developing countries for the acquisition and utilization of low to moderate cost hardware and software for resource management. It is the opinion of the editors that this

conference brought together a mix of professionals -- the providers of hardware and software and the users -- that needed to communicate with one another. The conference provided the proper setting for an exchange of information to take place, which will serve to improve the understanding of each other's needs in this emerging field.

We believe that what was started at the NASA-Ames Research Center in 1981 and expanded threefold in Baltimore by NASA-Goddard Space Flight Center's ERRSAC group will tremendously expand the domain and research activities of the specialists involved in resource management in the years to come.

As mentioned earlier, the conference proceedings are divided into two volumes. Volume I covers techniques, procedures and data bases while Volume II focuses on applications. Volume I is presented in two parts. Part 1 examines the techniques and procedures for extraction or reduction of data. Several papers compare techniques and procedures and data sources. Part 2 examines the process of integrating remotely sensed and other georeferenced data bases into geographic information systems for use in modeling and resource management applications. This is accomplished primarily by examining case studies and demonstration projects. Volume II is presented in five parts and is applications oriented. Part 1 examines the application of these systems to energy and environmental resource management. Part 2 describes the systems use in energy facility siting, while Part 3 examines its use in reclamation and surface mining. We also felt that this Volume would provide an appropriate framework for the examination of these systems use in various countries throughout the world. Several technique and procedure papers within an international context are also presented in Part 4. There were several

symposia and user forums presented throughout the conference and these are presented in Part 5. At the end of each Volume we have included the abstracts of the related poster sessions.

The wide spectrum of topics covered in these proceedings indicate that systems research techniques and procedures related to remotely sensed data and other georeference data bases for integration with geographic information systems are being increasingly applied to a larger number of resource management applications where they are helping us improve our ability to manage a finite set of the earth's resources in an environmentally sound manner.

In conclusion, the greatest contributions of these proceedings are that they help us create greater awareness of the issues among the designers and users of geographic information systems and georeferenced data bases. We hope that this will motivate each of us to devote more effort to this field and expand our interests even further in future meetings.

**Page intentionally left blank**

**Page intentionally left blank**

### ACKNOWLEDGEMENTS

I would like to express my thanks to all of the participants of the National Conference on Energy Resource Management for making this a successful conference and for submitting papers of excellent quality. I would also like to express my thanks to the members of the conference and paper review committees whose efforts contributed measurably to the success of the Conference.

My special thanks goes to Jim Brumfield of MURSAC, Bill Campbell of ERRSAC, Jimmie Weber of NASA--Headquarters, Rob Tanenhaus of IMI, and Germaine LaRouche of NRC, who were responsible for planning major segments of the Conference agenda. Bill, in particular, was responsible for planting the seeds from which this meeting grew. Many thanks to Dr. John McElroy who gave us the moral and fiscal support we needed at the early stages of conference planning. Thanks again to Jim Brumfield and Bill Campbell for their help in chairing the meeting. My special thanks also goes to Drs. Phil Cressy of ERRSAC, Dick Kott, and Ray Harris who helped us tie the various conference subthemes together in a coherent manner at our wrap-up session.

Many thanks to Chuck Killpack of IRIS International, Nicholas Short of NASA, and Rod Heller of Wirth Associates for their efforts in planning and conducting the pre-conference workshops.

A special thanks to the following chairpersons for their efforts in organizing a fine series of technical discussions: Jim Wray of USGS and George Jones of

Princess Coal--Use of Remotely Sensed Data and Geographic Information Systems for Energy Resource Management; William Campbell of ERRSAC and Ronald Ballard of NRC--Remote Sensing and Geographic Information Systems for Environmental Information Needs; Charles Cushwa of the Fish and Wildlife Service, Germain LaRouche of NRC, and Calvin DuBrock of the Pennsylvania Game Commission--Use of Remote Sensing for Environmental Analysis; Major Brian Lenz of the U.S. Air Force--Energy Management in the U.S. Air Force; Daniel Cotter and Russell Koffler of NOAA--Landsat Users Forum; Ray Harris, the Honorable General Secretary of the Remote Sensing Society and Joe Berry of Yale University--Use of Remote Sensing and Geographic Information Systems for Integration and Modeling; Brian Turner of ORSER, Scott Cox of NASA and John Estes of the University of California--Remote Sensing: Techniques and Procedures; Germain LaRouche of NRC and Ram Singh of the Department of Treasury--Use of Remote Sensing for Environmental Analysis; Robert Tanenhaus of Implementation Management International and Jimmie Weber of NASA Headquarters--Remote Sensing for Energy Resource Management, an International Perspective; Sue Norman and Dave Mouat of NASA--Energy Resource Management Data Bases; Rod Heller of Wirth Associates--Power Plant Siting; Herb Blodget and Marc Imhoff of NASA--Use of Remote Sensing for Reclamation and Surface Mining; and thanks to those who designed and conducted the poster sessions.

I am grateful for the support of the American Planning Association, U.S. Nuclear Regulatory Commission, NASA Goddard Space Flight Center--ERRSAC, and the U.S. Region of the Remote Sensing Society.

On behalf of the American Planning Association's Energy Planning Division, I would like to gratefully acknowledge the financial support of NASA Goddard

Space Flight Center--ERRSAC for the front end costs of this conference and the resources they are committing to the publication of these proceedings.

I would also like to thank the following exhibitors for the confidence they expressed in our conference and for their financial support: Earth Satellite Corporation of Chevy Chase, MD; U.S. National Oceanic and Atmospheric Administration of Washington, DC; Schiffman Energy Services, Inc. of Springfield, VA; U.S. Geological Survey of Reston, VA; Greenhorne and O'Mara, Inc. of Riverdale, MD; IRIS International, Inc. of Landover, MD; Computer Science Corporation of Silver Spring, MD; and Rogers, Golden and Halpern of Philadelphia, PA.

My gratitude to Diane Kugelman of the NASA Goddard Space Flight Center, ERRSAC who shepherded the various conference mailings and related data through the system so that we could maintain schedule.

Our thanks to the Baltimore Convention Bureau, and Trina Mello and Gary Flowers of the Baltimore Hilton for their fine efforts on our behalf.

My deepest gratitude goes to Monique Gormont of the Energy Planning Division who devoted many hours to all the small details and who handled the conference logistics, luncheons, audio visual aids, printing and directed the secretarial staff, and it is largely to her credit that all went so smoothly at the Conference. Thanks also to Pam Moore of APA who pitched in the last weeks to help us with the final details of the Conference.

November, 1982

Yale M. Schiffman  
Conference Chairperson  
President of APA Energy Planning  
Division  
Executive Director of the Center  
for Earth Resource Management  
Applications

Part 1

USE OF REMOTE SENSING AND GEOGRAPHIC INFORMATION

SYSTEMS FOR INTEGRATION AND MODELING



**Page intentionally left blank**

**Page intentionally left blank**

A STUDY OF FEATURE EXTRACTION  
USING DIVERGENCE ANALYSIS OF TEXTURE FEATURES

W. Hallada, B. Bly and R. Boyd  
Computer Sciences Corporation  
Silver Spring, Maryland 20910, U.S.A.

S. Cox  
NASA/GSFC, Code 902.1  
Greenbelt, Maryland 20771, U.S.A.

ABSTRACT

This paper presents an empirical study of texture analysis for feature extraction and classification of high spatial resolution remotely sensed imagery (10 meters) in terms of specific land cover types. Little is known as to which texture features are important for separating specific land covers with a per-pixel classifier. The principal method examined is the use of spatial gray tone dependence (SGTD). The SGTD method reduces the gray levels within a moving window into a two-dimensional spatial gray tone dependence matrix which can be interpreted as a probability matrix of gray tone pairs. Haralick et al (1973) used a number of information theory measures to extract texture features from these matrices, including angular second moment (inertia), correlation, entropy, homogeneity, and energy. The derivation of the SGTD matrix is a function of: 1) the number of gray tones in an image; 2) the angle along which the frequency of SGTD is calculated; 3) the size of the moving window; and 4) the distance between gray tone pairs. In this study, the first three parameters were varied and tested on a 10 meter resolution panchromatic image of Maryville, Tennessee using the five SGTD measures. A transformed divergence measure was used to determine the statistical separability between four land cover categories--forest, new residential, old residential, and industrial--for each variation in texture parameters.

1.0 INTRODUCTION

With the successful launch of Landsat 4, remote sensing investigators will be receiving multispectral imagery of land areas at more than one spatial resolution, 30 meters from the Thematic Mapper (TM) and 82 meters from the Multispectral Scanner (MSS). In the future, multispectral linear array (MLA) technology will provide digital imagery of even higher spatial resolution, on the order of 10 to 15 meters in the visible and near infrared (NIR), 30 meters in the short wave infrared (SWIR), and 120 meters in the thermal infrared [1]. It is apparent that more innovative approaches to digitally extract information from mixed resolution systems need to be examined. In terms of spatial and spectral resolutions, the method and data used for extracting information will obviously depend upon the application, the level of computing advancements, and the associated costs and benefits obtained by using digital data [2].

One analysis technique commonly used is supervised or unsupervised per-pixel multispectral classification. A problem investigators have found with this technique is that as spatial resolution increases, classification accuracies can decrease for land covers of high spatial complexity, such as encountered in urban and tropical environments. Markham and Townshend [3] found that when a higher percentage of mixed pixels exist, classification accuracies decrease. Conversely, spectral heterogeneity of other land cover classes tended to be averaged out at lower spatial resolutions. This resulted in less spectral overlap with other land cover classes, which in turn resulted in higher classification accuracies. Latty [4] found similar results for forest cover classification.

Higher spatial resolution therefore compounds the classification problem if the spectral information is not used in context with the spatial information. The classifier must be able to characterize the spatial context of spectral reflectances for each land cover type. It becomes readily apparent that this information needs to be incorporated into the classification process to make the digital extraction of information from future satellite imagery successful.

A number of algorithms and approaches have been developed to include spatial information in the classification process. Townshend and Justice [5] provide a brief summary of popular methods. These include texture analysis described by Haralick [6], spectral/spatial context used by Tilton and Swain [7], and categorical/spatial context used by Wharton [8]. The purpose of this paper is to examine one particular method of texture analysis introduced by Haralick et al [9] to extract spatial features that are described by second order statistics.

### 1.1 Texture Analysis

One common approach used to characterize an image's spatial information is to extract features for classification which measure the spatial arrangement of gray tones within a neighborhood of a pixel. This feature extraction method is referred to as texture analysis and includes a multitude of possible features that have been developed to describe image texture. Haralick [6] presents a complete literature review of texture analysis and Davis [10] presents some of the more recent developments. Connors and Harlow [11] investigated the mathematical and theoretical merits of various texture measures, whereas, Weszka et al [12] conducted an empirical comparison of various texture measures. Unfortunately, as noted by Townshend and Justice [5], more effort has been expended on the derivation of new texture measures than on evaluating the relative merits of each method for remotely sensed data. The use of texture analysis has been hindered with current satellite imagery because it effectively coarsens the spatial resolution. This introduces edge effects at the boundaries of land covers comprised of different gray tones or textures. Calculating a texture measure for a 5-by-5 window passed over a Landsat image coarsens the resolution to 400 meters. Future MLA resolutions should mitigate such effects.

Of the numerous texture measures available, the spatial gray tone dependence (SGTD) method has been used frequently by remote sensing investigators including Haralick et al [9], Haralick [13], Jensen [14], Jensen and Toll [15], Schowengerdt [16], and Weszka et al [12]. SGTD represents, both conceptually and computationally, an approach of greater breadth and complexity to texture extraction than such first order statistics as mean and standard deviation. The SGTD method transforms the gray values within a neighborhood of each pixel into a two-dimensional gray value co-occurrence matrix. This matrix  $P(i,j|\alpha,d)$  describes the frequency of occurrence of gray value pairs  $i, j$  separated by distance  $d$ , and angle  $\alpha$ , and therefore can be interpreted as a probability matrix of gray value pairs. Haralick et al [9] was the first to introduce a number of measures based on information theory to describe such matrices. Cox and Rose [17] developed computationally efficient software for calculating SGTD textures within the Interactive Digital Image Manipulation System (IDIMS). Measures implemented to date include: inertia (angular second moment difference); homogeneity (angular second moment inverse difference); correlation (covariance of neighboring pixels); entropy (average uncertainty of gray values); and, energy (angular second moment). These measures are mathematically summarized in Table 1, and are described in Table 2.

TABLE 1

- INERTIA (Angular Second Moment Difference)

$$INT = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i-j)^2 \frac{P(i,j|\alpha,d)}{\#R}$$

- ENERGY (Angular Second Moment)

$$ENG = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \left[ \frac{P(i,j|\alpha,d)}{\#R} \right]^2$$

- ENTROPY (Average Uncertainty of  $P(i,j|\alpha,d)$ )

$$ENT = - \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{P(i,j|\alpha,d)}{\#R} \text{Log} \left( \frac{P(i,j|\alpha,d)}{\#R} \right)$$

- HOMOGENEITY (Angular Second Moment Inverse Difference)

$$HOM = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{1}{1 + (i+j)^2} \frac{P(i,j|\alpha,d)}{\#R}$$

- CORRELATION (Covariance of Neighboring Points)

$$COR = \frac{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (ij) \frac{P(i,j|\alpha,d)}{\#R} - \mu_x \mu_y}{\sigma_x \sigma_y}$$

WHERE #R = Number of Neighboring Cells

$Ng$  = Number of Gray Tones, and  $\mu_x, \mu_y, \sigma_x, \sigma_y$  are the means and standard deviations of the marginal distributions associated with  $P(i,j|\alpha,d)/\#R$ .

TABLE 2

- INERTIA
  - Measures tendency to concentrate probability away from the main diagonal of the co-occurrence matrix.
  - Related to gray value variance.
  - Inversely proportional to image coarseness, or contrast.
  - Lower bound when texture is entirely monotone.
- HOMOGENEITY
  - Measures the similarity of neighboring pixels.
  - Flat textures will give higher values.
  - Upper bound when all probability lies on the main diagonal of the co-occurrence matrix.
- CORRELATION
  - Measures the covariance of neighboring pixels.
  - Zero when all pixels are independent.
  - Natural scenes tend to have lower values.
  - Has the largest values for periodic patterns.
- ENTROPY
  - Measures the average uncertainty of gray values pairs.
  - Upper bound when all probabilities are equal.
  - Lower bound when one gray tone pair has a probability of 1.
  - Invariant to monotonic gray tone transformations.
- ENERGY
  - Measures the average certainty of occurrence of gray value pairs.
  - Lower bound when all probabilities are equal.
  - Upper bound when only one probability appears.
  - Homogeneous areas have higher energy.
  - Invariant to monotonic gray value transformations.

In summary, the derivation of the SGTD matrix is a function of the following parameters:

1. The number of gray levels within an image. The computation of the texture feature is related to the square of the number of gray levels.
2. The angle along which the frequency of occurrence is derived. For example, there are four independent angles for a distance of one, and eight for a

distance of two resulting in four and eight independent features for each image. Haralick et al [9] suggested using the average and range (which are invariant under rotation) as inputs to the classifier.

3. The size of the moving window. Small window sizes will not adequately sample the SGTD probabilities of land cover classes [18]. Conversely larger window sizes will degrade the resolution of remotely sensed imagery.
4. The distance between pixels in tabulating the co-occurrence matrix. Haralick [6] argued that the co-occurrence matrix for a single distance contains most of the significant texture information.

An empirical investigation of the effects of quantization level, orientation, and window size upon classification accuracy of remotely sensed imagery does not, to our knowledge, exist in the literature. Little is known about the effectiveness of texture analysis in terms of sensor spatial resolution and the spatial frequencies of land cover on the ground. In addition to the above parameters, approximately two dozen dependent SGTD features can be used [6]. Compression techniques using eigenvector analysis were proposed by Tou and Chang [19] to reduce this large dimensionality for features comprised of SGTD measures from different angles and distances.

## 2.0 STUDY SITE AND TEST DATA

For the texture investigations, a test site containing a mixture of urban, forest, and agricultural land covers was chosen in order to provide a variety of textures to study. The digital imagery was acquired using a Daedalus DS-1260 MSS flown on April 7, 1977 over Maryville, Tennessee. Flown from an altitude of 3,000 meters, the instantaneous field of view (IFOV) at nadir was 8.25 meters. This data set was one of many processed by Geospectra Corporation of Ann Arbor, Michigan under contract to the National Oceanic and Atmospheric Administration to provide multispectral imagery of diverse land cover types. In processing the data, Geospectra Corporation resampled the original scanner data to 10 meter resolution and rectified it to a Universal Transverse Mercator projection. Additional processing included interband averaging to simulate Thematic Mapper bands, and contrast stretching. The contrast stretching, interband averaging, resampling, and degradation methods made the utility of the data questionable [20]. For future studies it is suggested that the methods reported herein be attempted with data more representative of future MLA satellite instruments.

The study site, Maryville, Tennessee is a small city of 14,000 people located 10 miles south of Knoxville. The entire test site includes a range of residential densities, commercial and industrial areas including infrastructure such as roads and airports, forested areas and agricultural fields. The entire test image covers an area of approximately 5 square kilometers, with the street pattern oriented at 45 degrees off the image line and sample axes. Visual interpretation of this data, clearly reveals that applications of remote sensing for urban studies would readily benefit from a 10 meter MLA instrument.

### 3.0 METHODS

To reduce the number of features and preserve as much spatial information as possible, a panchromatic image was synthesized from the green (.55-.60  $\mu\text{m}$ ), red (.6-.69  $\mu\text{m}$ ) and near-infrared (.8-1.1  $\mu\text{m}$ ) bands. The norm, or the square root of the sum of the three squared gray values was used to simulate a panchromatic image. The correlation of the panchromatic band with the green, red, and near-infrared bands was .89, .93, and .82, respectively. The resulting image is shown in Figure 1.

#### 3.1 Test Site Selection

From this image four training sites were chosen to study the effect of window size, quantization, orientation, and SGTD measure upon classification accuracy. Each test site was selected based on land cover, visual texture, and size. The four sites were: 1) mature deciduous forest; 2) old residential composed of mature deciduous trees, old homes and narrow paved roads; 3) new residential composed of large lots, larger ranch style homes, wider roads and few trees; and, 4) an industrial site with concrete parking lots and large buildings with linear shadows. Roads in both residential sites were oriented at 45 degrees and the buildings in the industrial site horizontally. The four unique complex land covers, shown in Figure 1, provided a good basis for comparing texture features.

Each test site consisted of a 40-by-40 pixel block. After applying the five texture measures on a pixel by pixel basis, statistics were calculated for a 20-by-20 pixel training block centered within each test site to eliminate any harmful effects from edges between sites, as well as to provide a large sample of 400 pixels per land cover.

#### 3.2 Feature Extraction

The effects of quantization, window size, and orientation angle were extensively tested using the inertia texture measure (Table 1). Only the window size parameter was varied for the other four texture measures of energy, entropy, correlation, and homogeneity. The quantization level was tested by requantizing the gray levels from 256 levels to 4, 8, 16, 32, 64 and 128 gray levels using a simple linear mapping. Spatial gray tone inertia features were then generated using sliding windows of 3-by-3 (30 meter textures) to 13-by-13 (130 meter textures), in increments of two, for the four possible orientations (0, 45, 90 and 135 degrees) at a distance of one. A total of 168 texture features were therefore created and evaluated; that is, 4 orientations, 6 window sizes, and 7 quantization levels. The means and covariances of the four orientations combined with the original gray tone image were calculated for each quantization level, window size, and training site.

### 3.3 Transformed Divergence Analysis

Rather than perform an actual classification for various combinations of texture features followed by an accuracy test using test and training sites, a statistical measure of separability was employed as a predictor of classifier performance. Once the statistics were calculated for each feature set combination, a transformed divergence measure was used to determine the inter-class separability of the four land covers.

Divergence [21] between class pairs  $i$  and  $j$  is defined as:

$$D_{ij} = 1/2 \operatorname{tr} \left[ (\Sigma_i - \Sigma_j) (\Sigma_j^{-1} - \Sigma_i^{-1}) \right] + \\ 1/2 \operatorname{tr} \left[ (\Sigma_i^{-1} + \Sigma_j^{-1}) (M_i - M_j) (M_i - M_j)^T \right]$$

WHERE  $\Sigma$  = Class covariance matrix

$M$  = Vector of class means

$\operatorname{tr}$  = trace (sum of the diagonal elements).

Because divergence increases without bound as statistical separability between classes increases, Swain and Davis [21] defined a saturation transform which provides a measure more closely corresponding to percent correct classification. The transformed divergence expression is:

$$TD_{ij} = 2,000 \left[ 1 - \exp (-D_{ij}/8) \right].$$

This measure has a saturating behavior, that is, percent correct classification saturates at 100 percent when a certain level of statistical separability is reached ( $TD = 2,000$ ).

There are some disadvantages in using transformed divergence as a measure of statistical separability between class pairs. For example, two class densities having equal mean vectors but non-equivalent covariance matrices may result in a transformed divergence of zero [22]. Furthermore, there is no estimate for a lower confidence limit for the relation between transformed divergence and percent correct classification. In lieu of alternative measures, transformed divergence is very efficient computationally, and affords a relative measure of performance without doing an actual classification.

### 4.0 RESULTS

The average transformed divergences (TD) of all land cover pairs are plotted in Figure 2 for each window size of inertia calculated from data with 128 gray levels. As the window size increased, the average TD value increased. Combining the four orientations into a single normalized measure significantly reduced the average TD. The addition of the gray tone image increased separability but not enough for acceptable classification accuracy. The increases in average TD of the four orientations behaved in a logarithmic fashion and began to level off at a window size of 11 pixels (110 meters).



Figure 3 plots the TD values using four orientations of inertia for each land cover pair except for those having forest. The TD between forest and all other land covers was 2000. Apparent length of the lines connecting different window sizes in Figure 3 is proportional to the added separability resulting from an increase in spatial information. Four orientations of inertia should, therefore, provide features which may be used to classify forest with 100 percent accuracy at any window size. As expected, the separability between the two residential classes was the lowest for all window sizes. A larger window size may be needed to improve the separability between these two categories.

The effect of gray level quantization upon TD is shown in Table 3 for the two residential categories. A decrease in separability accompanied by a decrease in gray level quantization does not occur until approximately level 16. At this level the separability between the two land covers decreases somewhat. A larger window size did compensate for this decrease.

TABLE 3

Transformed divergence between new and old residential  
for changes in window size and gray level quantization.

Quantization Level	Transformed Divergence Values					
	3x3	5x5	7x7	9x9	11x11	13x13
4	184	413	668	1120	1430	1627
8	251	558	874	1325	1584	1699
16	238	519	822	1266	1638	1797
32	265	563	875	1319	1667	1800
64	256	553	869	1319	1657	1796
128	256	549	867	1323	1672	1801
256	258	556	874	1327	1675	1806

Quantization level may affect both the feature extraction and classification process, and hence affect accuracies. As noted earlier, reducing the number of gray tones in the input image will decrease the size of the SGTD matrix as well as reduce computation time. Furthermore, most maximum likelihood classification software in image processing systems are written to classify byte data with 256 gray levels. The result of a texture transform is a real number which must be scaled and quantized to fit within 256 gray levels.

Table 4 shows the result of requantizing the four orientations of inertia for new and old residential. Similarly, there was a reduction in separability due to the requantization of the texture measure, but the reduction does not seem significant.

TABLE 4		
Transformed Divergence Between Old and New Residential.		
<u>Window Size</u>	<u>256 Levels</u>	<u>Floating Point</u>
7x7	736	875
13x13	1710	1800

Additional insight into the effect of orientation on texture feature extraction was gained through plotting transformed divergences for subsets of 2 and 3 orientations, as shown in Figure 4. From this figure, it is evident that all four orientations were important for separating various urban land covers. No subset of 2 or 3 orientations provided adequate separation between all land covers, although larger window size may compensate for the loss of orientation features.

The difference in separability due to the type of SGTD measure is shown in Figure 5. Homogeneity was better at separating the industrial and residential land covers, whereas inertia provided the best feature for separating forest from all other land cover categories. The average transformed divergence plotted in Figure 6 indicates that inertia had the best overall separability performance for separating the four land cover types, and correlation was the worst.

## 5.0 CONCLUSIONS

The results obtained in this empirical study demonstrated that quantization level, window size, and orientation are very important parameters to consider when using the SGTD method for extracting texture features from high spatial resolution remotely sensed imagery. Although transformed divergence did not provide a perfect measure of classification accuracy; it did provide a robust method to evaluate texture features.

In summary, the following results were found:

- 1) As window size increased, class separability increased logarithmically. Separability between certain land covers was maximized at smaller window sizes depending upon the SGTD measure.

- 2) Class separability was very sensitive to SGTD orientation. A subset of orientations as well as the norm of the four orientations proved inadequate for separating the four land covers.
- 3) Class separabilities did not begin to decrease with decreasing gray levels until 16 gray levels. At 16 or 8 gray levels larger window sizes were needed to preserve separability.
- 4) Rescaling texture features from a 32-bit real number into 256 gray levels reduced class separability; however, this did not seem as important a factor as window size and orientation.
- 5) Of the five SGTD measures tested, inertia had the best performance for separating the few land covers investigated herein.

## 6.0 RECOMMENDATIONS

Quantitative criteria should be developed to determine the spatial resolutions optimal for using texture analysis for specific applications, e.g. urban remote sensing. Work similar to that reported herein should be attempted with remotely sensed data of various spatial and spectral resolutions, and to include further empirical comparisons of other texture features. These feature combinations should encompass first order statistics [6] and the recently developed texture energy measures [10]. Additionally, it is recommended that algorithms, which process image data spatially be implemented on parallel processors such as the massively parallel processor [23].

Furthermore, it is suggested that recent developments in image texture analysis, as reported in the pattern recognition literature, be attempted with remotely sensed imagery. Such recent developments are: the use of the co-occurrence matrix directly in a classification algorithm [18, 24]; the use of segmentation techniques to partition remotely sensed imagery into unique texture regions, such that boundaries between regions are correctly represented [10, 25]; and, research into multispectral texture models and classification [26].

## 7.0 REFERENCES

1. Mika, A.M., 1982. Design tradeoffs for a multispectral linear array (MLA) instrument. In the Multispectral Imagery Science Working Groups: Final Report, Vol. 3, Contributed Papers. S. Cox, Ed. September 1982. NASA Conference Publication in Press.
2. Welch, R., 1982. Spatial resolution requirements for urban studies. International Journal of Remote Sensing. 3: 139-146.
3. Markham, B.L. and Townshend, J.R.G., 1981. Land cover classification accuracy as a function of sensor spatial resolution. Proceedings of the 15th International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan (in press).

4. Latty, R.S., 1981. Computer based forest cover classification using multispectral scanner data of different spatial resolutions. LARS Technical Report 052081. Laboratory for Applications of Remote Sensing, West Lafayette, Indiana, U.S.A.
5. Townshend, J. and Justice, C., 1981. Information extraction from remotely sensed data. A user view. International Journal of Remote Sensing. 2: 313-329.
6. Haralick R.M., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE. 67: 786-804.
7. Tilton, J.C. and Swain, P.H., 1981. Incorporating spatial context into spatial classification of multidimensional image data. LARS Technical Report 072981. Laboratory for Applications of Remote Sensing, West Lafayette, Indiana, U.S.A.
8. Wharton, S.W., 1982. A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. Pattern Recognition. 15:317-324.
9. Haralick, R.M., Shanmugam, K., and Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics. SMC-3:610-621.
10. Davis, L.S., 1982. Image texture analysis: Recent developments. In Proceedings of 1982 IEEE Conference on Pattern Recognition and Image Processing, Las Vegas, Nevada, pp. 214-217.
11. Conners, R.W., and Harlow, C.A., 1980. A theoretical comparison of texture algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-2:204-222.
12. Weszka, J.S., Dyer, C.R., and Rosenfeld, A., 1976. A comparative study of texture measures for terrain classification. IEEE Transactions on Systems, Man and Cybernetics. SMC-6:269-285.
13. Haralick, R.M., 1974. Combined spectral and spatial processing of ERTS imagery data. Journal of Remote Sensing of the Environment. 3:3-13.
14. Jensen, J.R., 1979. Spectral and textural features to classify elusive land cover at the urban fringe. The Professional Geographer. 31:400-409.
15. Jensen, J.R. and Toll, D.L. 1982. Detecting residential land-use developments at the urban fringe. Photogrammetric Engineering and Remote Sensing. 48:629-643.
16. Schowengerdt, R.A., 1981. Texture feature extraction. Final Report, USGS Grant #14-08-0001-G-664. Applied Remote Sensing Program, University of Arizona, Tucson, Arizona, U.S.A.

17. Cox, S. and Rose, J., 1982. Texture functions in image analysis, computationally efficient solution. In press.
18. Vickers, A.L. and Modestino, J.W., 1982. A maximum likelihood approach to texture classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-4:61-68.
19. Tou, J.T. and Chang, Y.S., 1977. Picture understanding by machine via textural feature extraction. In Proceedings 1977 IEEE Conference on Pattern Recognition and Image Processing. (Troy, New York), pp. 392-99.
20. Communication with EROS Data Center, May, 1982.
21. Swain, P.H. and Davis, S.M., 1978. Remote Sensing: The Quantitative Approach, New York: McGraw-Hill.
22. Swain, P.H., Robertson, T.V., Wacker, A.G., 1971. Comparison of the divergence and B-distance in feature selection. LARS Information Note 200871. Laboratory for Applications of Remote Sensing, West Lafayette, Indiana, U.S.A.
23. Strong, J.P., Schaeffer, D.H., Fischer, J.R., Wallgren, F.A., and Bracken, P.A., 1978. The massively parallel processor and its applications. Proceedings of the 12th International Symposium on Remote Sensing of Environment, ERIM, University of Michigan, Ann Arbor, pp. 37-45.
24. Zucker, S.W. and Terzopoulos, D., 1982. Finding structure in co-occurrence matrices for texture analysis. In Image Modeling. A. Rosenfeld, ed. New York: Academic Press, pp. 423-445.
25. Davis, L.S. and Mitichie, A., 1982. MITES: A model-driven, iterative texture segmentation algorithm. Computer Graphics and Image Processing. 19:95-110.
26. Rosenfeld, A., Wang, C., and Wu, A.Y. Multispectral texture. IEEE Transactions on Systems, Man, and Cybernetics. SMC-12: 79-84.

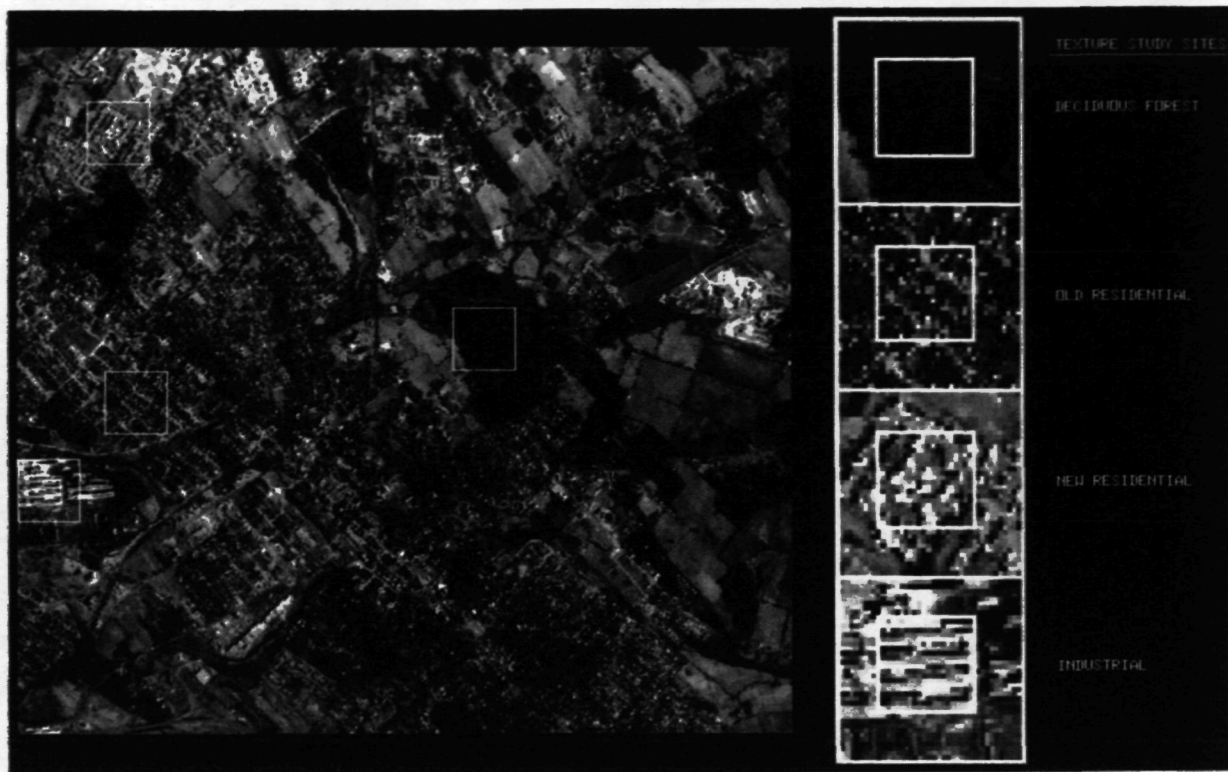


Figure 1. Synthesized panchromatic image and texture study sites.

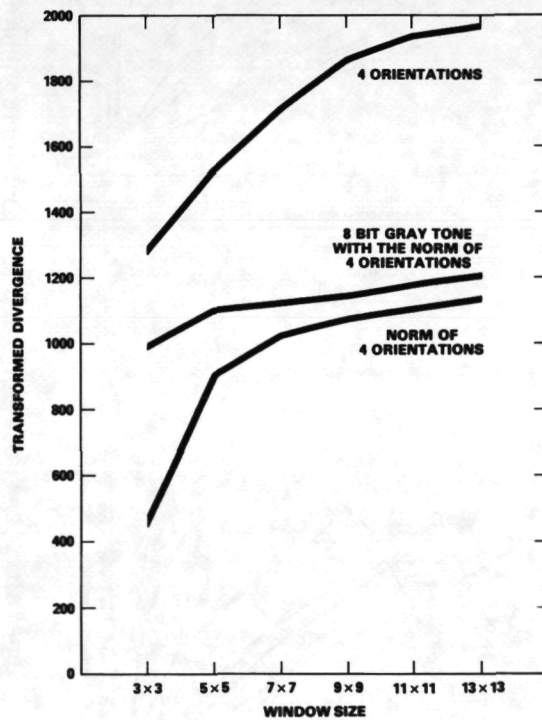


Figure 2. Average separability using 4 orientations of inertia, versus the norm and norm plus panchromatic image.

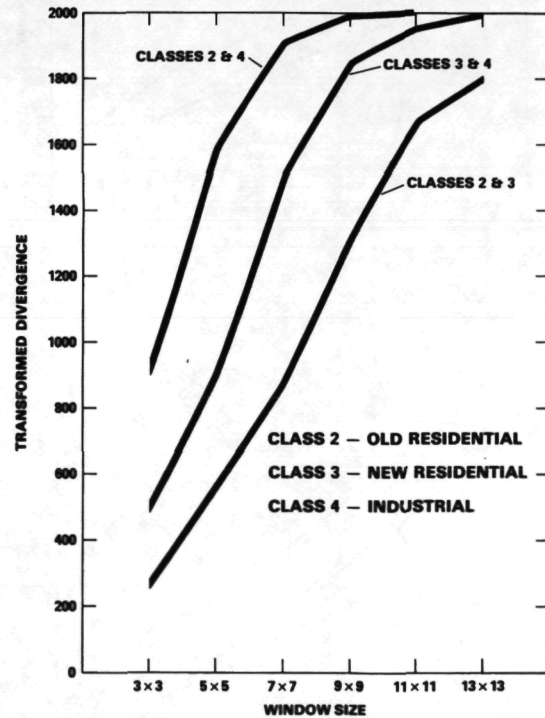


Figure 3. Separability between land covers using 4 orientations of inertia.

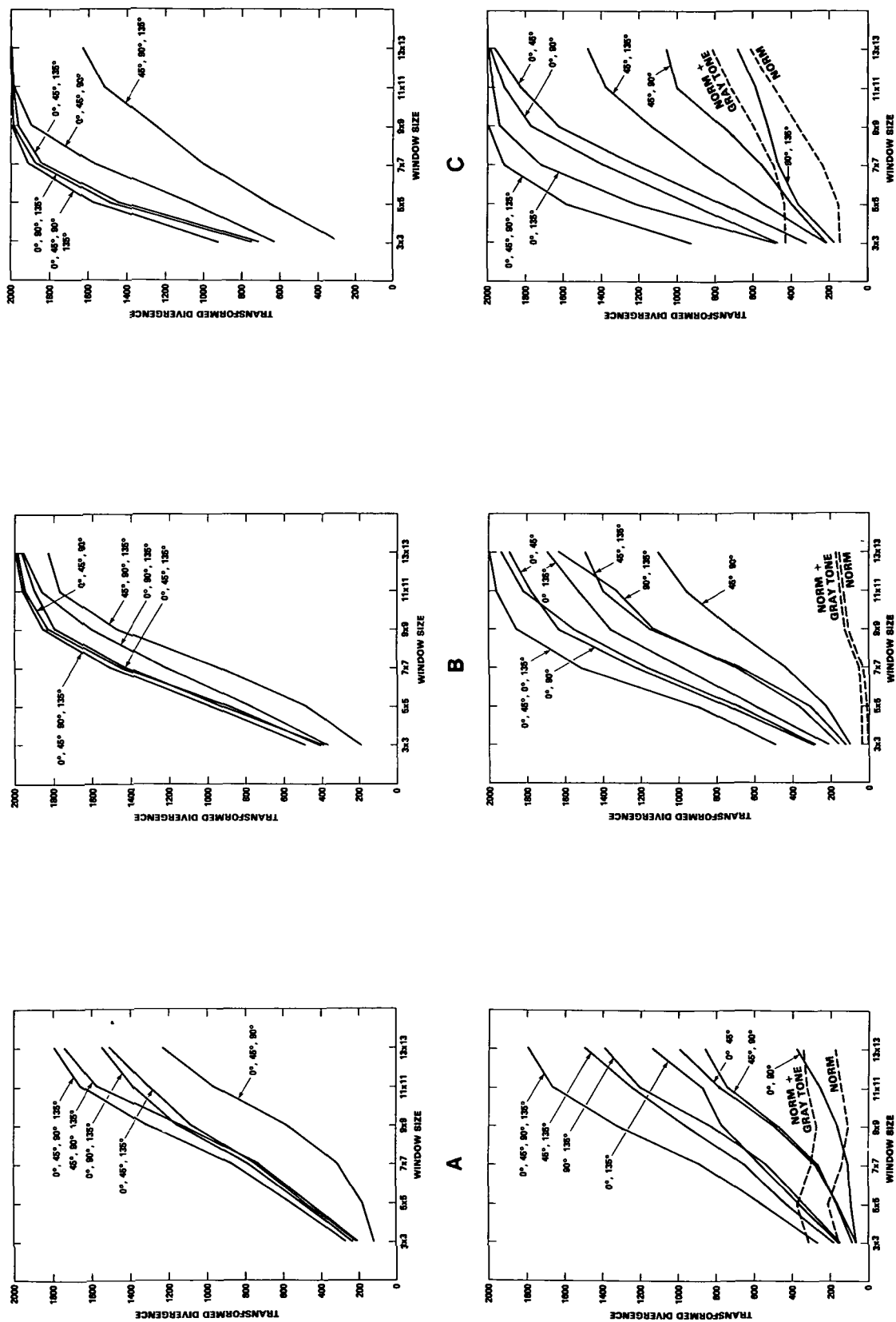


Figure 4. Separability between land covers using spatial gray tone inertia, a) new and old residential, b) new residential and industrial, and c) old residential and industrial.

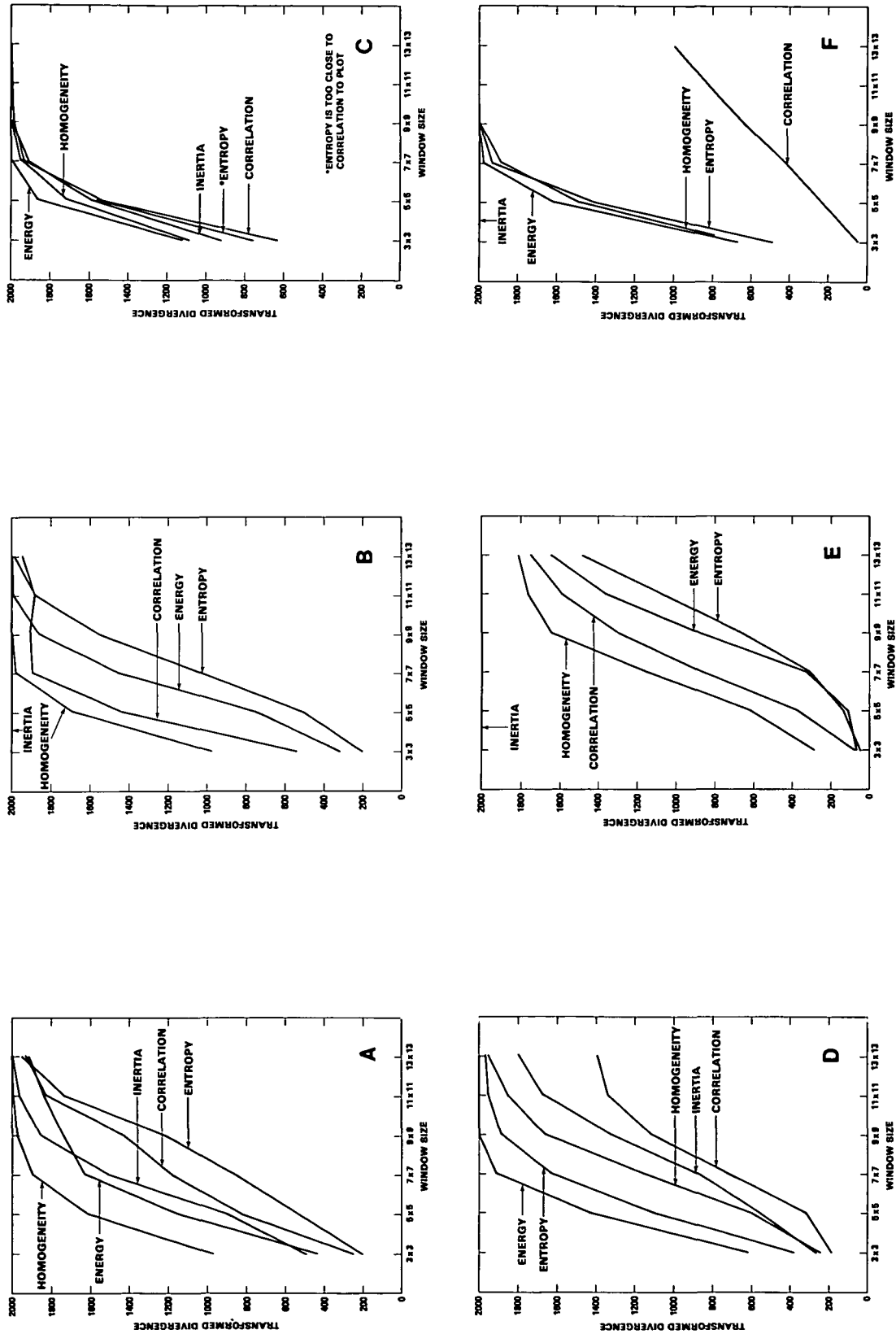


Figure 5. Separability between a) new residential and industrial, b) industrial and forest, c) old residential and industrial, d) new and old residential, e) new residential and forest, and f) old residential and forest.



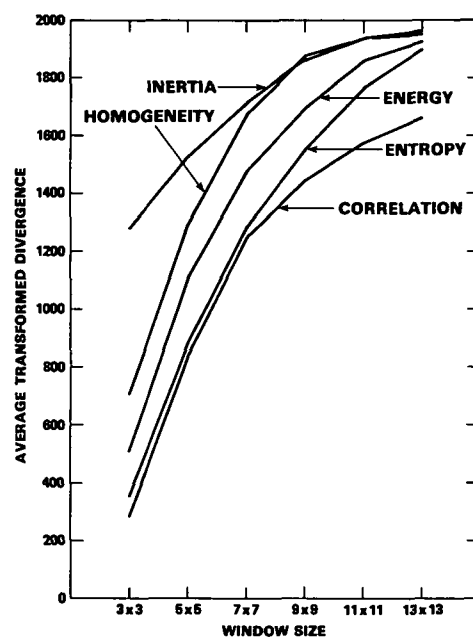


Figure 6. Average interclass separability between land covers using 4 orientations of SGTD measures.

# PRINCIPAL COMPONENTS AS A DATA REDUCTION AND NOISE REDUCTION TECHNIQUE

MARC L. IMHOFF

WILLIAM J. CAMPBELL

NASA/Goddard Space Flight Center  
Greenbelt, Maryland 20771

## ABSTRACT

The objectives of this study were to: (1) Assess the potential of principal components as a pipeline data reduction technique for thematic mapper data, and (2) Examine principal components analysis and its transformation as a noise reduction technique.

Two primary factors were considered:

1. How might data reduction and noise reduction using the principal components transformation affect the extraction of accurate spectral classifications, and
2. What are the real savings in terms of computer processing and storage costs of using reduced data over the full 7-band TM complement?

An area in central Pennsylvania was chosen for a study area. The image data for the project were collected using the Earth Resources Laboratory's Thematic Mapper Simulator (TMS) instrument. The TMS records data in seven band widths (.46-.52, .53-.60, .63-.69, .77-.90, 1.53-1.72, 2.04-2.24, and 10.43-12.33  $\mu\text{m}$ ) with a ground instantaneous field of view (GIFOV) of 30 meters. A set of surface feature verification sites corresponding to desired land cover/land use classes were geodetically measured and photographed using field teams and low altitude color infrared aerial photography. The photographs with the surface verification site boundaries were digitized, registered, and merged with the TMS data. A percentage of the surface feature verification sites was used for spectral signature training while the remaining sites were utilized for accuracy assessment.

A principal components analysis and its associated transformation was applied to six of the seven spectral bands. The thermal band 7 was not included in the initial transformation. Cost and classification accuracy comparisons were made using a supervised classification procedure applied to selected subsets of the transformed data and compared with results obtained by applying the same procedure to the full 6 band complement.

Classifications were made on a subset consisting of three principal components axes and the full 6 band contingent for comparison. Overall classification accuracy for the transformed and reduced data was down 4 percent from that achieved using the full 6 bands. Processing costs for the transformed and reduced data were less than 53 percent of the costs required to process the 6 band data.

## INTRODUCTION

The Thematic Mapper (TM) instrument on Landsat-4 is collecting nearly 15 times more data per unit area than the Multispectral Scanner (MSS). As a result of this increased data volume, data reduction techniques may be desired, or even required, by some users to reduce cost impacts in computer processing and personnel time.

Three basic techniques that are commonly used to affect the reduction of digital satellite data are: (1) Band selection, (2) Data transformations based on preliminary spectral classification, i.e., canonical analysis, and (3) data transformations based on overall data statistics, i.e., principal components.

Recent studies using band selection techniques applied to Thematic Mapper Simulator (TMS) data have yielded mixed results. Dottavio and Williams (1982) found that a band subset of three carefully selected channels slightly improved classification accuracies over those attained using the full compliment for forest cover types in North Carolina. Gervin et al. (1982), however, found that using all bands yielded higher classification accuracies than band subsets for mapping cover types in Michigan. Dottavio and Williams used the NS-001/MS Thematic Mapper Simulator and Gervin used the 18ML Scanner (also used in this study). The difference in their results may be due to the differences in the instruments themselves.

Data reduction techniques using statistical transformations have also been explored. Canonical analysis is known to improve class separability for most of the classes input to the transformation but the separability of some classes are sacrificed for the increased separability of other classes. At this writing it is not well documented as to exactly how the transform affects overall classification accuracy in different situations and objectives.

Canonical analysis also requires a preliminary spectral classification in order to develop the transformation matrix, a factor that makes this technique actually more expensive in terms of processing costs than simply classifying the full band contingent (Imhoff and Petersen, 1980).

Principal components can be used to exert a mathematical transformation requiring little preliminary spectral analysis prior to its application. Care must be exercised to include all important target features in the initial development of the transformation matrix to achieve an optimal result, but this requires only a fraction of the effort required for creating a preliminary spectral classification.

Principal components therefore appears to be a well suited technique for quick, inexpensive data reduction. The question remains, however, as to how many transformed data channels can be removed without reducing classification accuracy below tolerable limits. It also remains to be determined what those tolerable accuracy limits are in relation to the limitations of data processing and analysis costs. This is most probably a variable that will change with each user and application.

## OBJECTIVE

The objective of this paper was to examine the potential of principal components analysis and its transformation as a pipeline data reduction and noise reduction technique. The two criteria by which the success of this technique was assessed were: (1) classification accuracy, and (2) data processing and analysis costs.

## MATERIALS AND METHODS

The Thematic Mapper Simulator (TMS) data used in this study was acquired using the 18ML Scanner and is composed of seven bands (0.45-0.52, 0.52-0.60, 0.63-0.69, 0.76-0.90, 1.55-1.75, 2.08-2.35, 10.4-12.5  $\mu\text{m}$ ) (ORI, 1982). The instrument has a GIFOV of 30 x 30m and the data were collected from an airborne Learjet aircraft at an altitude of 45,000 feet. The area chosen for analysis was a site surrounding the Susquehanna nuclear steam electric generating facility in Berwick County, Pennsylvania (figure 1). This project was carried out as part of the overall NASA/NRC Energy Facility Siting Program (Campbell, 1982).

The TMS data suffered from several problems:

- a. Considerable image distortions due to aircraft flight path movement were apparent in all bands,
- b. Calibration problems and electronic noise appeared in the imagery in the form of line striping and beat patterns, and
- c. A high frequency spatial distortion possibly due to aircraft and/or instrument jitter was present in all bands.

The distortions caused by aircraft/scanner jitter were not immediately removable and left unchanged. The effects of problems a and b above were handled as described below.

Prior to collecting training site statistics some extensive pre-processing was undertaken to remove some of the radiometric and geometric distortions inherent in the aircraft-collected TMS data. Two primary steps were taken:

- a. Radiometric adjustment for scan angle effects. The look angle and subsequently the atmospheric path lengths vary systematically as the TMS scans across the flight path. As a result, reflectance data recorded for a particular target feature near nadir appear different from that of the same target feature off nadir with a longer atmospheric path length. This factor causes confusion in the classification of land cover categories over the image. In order to compensate for this effect, the raw TMS data were normalized to a predicted radiometric response at nadir.



22

b. Geometric correction. Geometric distortions perpendicular to the flight line were also inherent in the TMS data. These distortions were caused by variations in the look angle during data collection. A control grid was developed with a cell size equal to the TMS resolution cell size at nadir. A nearest neighbor resampling algorithm was then used to fit all of the TMS image data to the control grid.

#### GROUND TRUTH

Once the primary geometric and radiometric distortions were removed, the TMS data were geodetically precision registered to a series of digitally encoded, low altitude color IR aerial photographs which were in turn geodetically registered to US Geological Survey (USGS) 7.5 minute topographic quadrangles.

In order to exercise scientific control in comparing the classification accuracies achieved for the unaltered and transformed data sets, precise ground truth data were collected. The method that was designed for the study was to collect ground truth data coincident with the low altitude overflights. In reality, scheduling the concurrence of these events with cloudless, clear weather proved to be an impossible task. However, the time interval between the three events was 6 weeks, not optimum but adequate.

A rigorous cluster sampling procedure was designed to combine the ground truth surveys with the low altitude color IR digitized photography. Areas were randomly selected from USGS 7.5 minute quadrangle maps covering each test site. The randomly selected sites were visited and photographed in color and color IR. A professional survey team provided locational accuracy to within  $\pm 1$  foot with a laser geodimeter. The survey data were then combined with the digitized color IR data which were digitally registered with the USGS 7.5 minute quadrangle maps to produce georeferenced ground truth which were in turn used to generate pixels of known identity in the areas sampled for both training set generation and accuracy assessment. The main advantage of this procedure is that cluster sampling provides identities for more pixels per area visited than systematic sampling or simple random sampling.

Using the cluster sampling and survey technique 180 training sites were documented and registered. Approximately half of the training sites were used for signature derivation and the remainder were reserved for classification accuracy assessment.

#### DATA REDUCTION--PRINCIPAL COMPONENTS

In order to effect a data reduction, a general principal components analysis and its transformation was used to create a new set of data channels whereby more of the system variance might be explained by a fewer number of data channels or axes.

Principal components analysis and its transformation was selected due to its relative simplicity and general availability. Jet Propulsion Laboratory's VICAR, ESL's IDIMS, and the Pennsylvania State University's ORSER system all have principal components options. Principal components is a technique whereby a new set of axes is defined for the data such that the first principal component or axis explains as much of the total variance as can be explained by any single variable or axis. The second principal component or axis explains as much of the remaining variance as can be explained by any axis or orthogonal (uncorrelated) to the **first**. The third principal component continues this process and so on until the dimensionality of the data is exhausted (Merembeck and Borden, 1978). The effect is that most of the information inherent in the many spectral bands is combined or explained by one, two, or three of the principal components.

In this application a general principal components (PC) analysis was used to derive the transformation matrix for the TMS data. A simple polygon targeting training site selection and statistical calculation program was used to determine mean responses for each band and a variance-covariance matrix for a general cross section of the data. The transformation training site transected all of the major cover types and/or target features found in the data. A set of Eigen values was calculated from the training statistics and the transformation matrix was generated.

Once the transformation matrix was applied, a variance-covariance matrix and correlation matrix was generated to determine the effectiveness of the transform and compare the new axes with the unaltered data (figure 2). In this case the data represented by axes 1-3 accounted for 98 percent of the total system variance and raw channels 2, 3 and 4 had the highest correlations with the first three principal axes. For this application the data represented by axes 1, 2 and 3 were used for classification purposes for comparison with the full 6-band contingent. The data represented by axes 4, 5, and 6 were discarded.

For the purpose of simplicity, throughout the remainder of the text, the full 6-band unaltered data will be referred to as "raw" or "raw 6-band" data and the transformed and reduced data will be referred to as "PC" or "PC 3 axes" data.

#### ACCURACY ASSESSMENT

As described above, the training site boundaries were delineated on paper copies of the low altitude aerial photography. These boundaries were then transferred to the digitized version of the same photography using an interactive CRT and track ball-driven cursor.

Once in digital format the randomly selected training site boundaries for each target feature were divided in two categories:

CORRELATION MATRIX PRINCIPAL COMPONENTS AXES 1 2 & 3  
VS  
UNALTERED 6 CHANNEL DATA

UNALTERED CHANNELS	PRINCIPAL COMPONENTS AXES		
	1	2	3
1	-.72	.65	.27
2	-.67	.79	.20
3	-.72	.80	.25
4	.91	-.10	-.69
5	.23	.55	-.35
6	-.35	.65	-.10

COVARIANCE MATRIX FOR PRINCIPAL COMPONENTS TRANSFORMED DATA  
6 CHANNEL

PC AXES	1	2	3	4	5	6
1	526.51					
2	-74.02	310.24				
3	15.11	18.05	48.10			
4	10.36	7.63	-1.22	9.98		
5	5.99	1.45	1.71	-.12	6.48	
6	-5.41	.28	-.83	.19	-.13	3.62

FIGURE 2. VARIANCE-COVARIANCE MATRIX AND CORELATION MATRIX  
FOR 6 BAND AND PRINCIPAL COMPONENTS TRANSFORMED DATA



a. A statistical (STATS) category from which spectral signatures for classification were developed, and

b. An accuracy assessment category (ACC) against which the classification was to be tested.

The two sets of training site boundaries or polygons were stored as images, each polygon retaining its geometry and spatial juxtaposition as it appeared on the georegistered digital data. The two sets of polygon images were then used for comparison with classified data. The STATS polygon boundaries were transferred to both the PC TMS data and the raw TMS data for the generation of spectral signature statistics for classification. Once the spectral signature banks had been developed for the PC and raw TMS data, the two scenes were classified using a maximum likelihood classifier. The same algorithm was used to generate the spectral signatures and classify both the PC 3 axes and raw 6-band data sets.

After classification, accuracy assessment was made by creating contingency tables of the classified PC and raw TMS data sets against the ACC image. Calculations derived from the contingency tables provided classification accuracy statistics in the form of:

- a. Probability that a pixel classified as class i is class i,
- b. probability that a pixel that is class i is classified as class i, and
- c. overall (combined for all classes) probability of correctly classifying a pixel given this set of circumstances.

#### COST ASSESSMENT

Cost assessment was made by documenting the time required to generate the classifications for both the raw 6-band and PC 3-axes data sets. The items measured were:

- a. Central Processing Unit (CPU) time (seconds)
- b. Computer connect time (minutes)
- c. Man-hours

The costs were documented in the form of time and not dollars since the dollar/time relationship changes for each user and set of circumstances.

The time required for the principal components analysis and transformation was included in the costing of PC 3-axes classifications. The time costs of preprocessing were common to both data sets and not included.

## RESULTS

The contingency tables comparing the classified data sets with the ACC image or accuracy data set provided information concerning the classification accuracies for each data set. Overall statistics and statistics for each class were calculated from the contingency tables comparing classification accuracies or classification performance level for the raw 6-band data and the PC 3-axes data (figure 3). In general, the accuracy statistics were fairly good.

Accuracies for some classes such as coniferous forest, orchards, mixed forest and meadow were low due to the lack of good training sites for these cover types in the Berwick area.

The classification performance comparison revealed that the overall probability of correctly classifying a pixel was slightly lower for the PC 3-axes data (4.3 percent) than for the 6-band raw data (figure 4). A class by class analysis reveals that for most classes the probability that a pixel classified as class i is class i and the probability that a pixel in class i is classified as such both decreased slightly using data reduction. The probabilities of correct classification for a few classes, however (barren, meadow, and water), actually increased using the reduced data.

The cost analysis calculated the central processing unit (CPU) time (in seconds), the man-hours, and the computer connect time required for the generation of training site statistics and the actual classification of the raw 6-band and PC 3-axes data sets. The time required for the TMS preprocessing was not included as it was the same in both cases. The cost statistics for generating the PC analysis and transformation were included in the costing of the transformed data.

The cost comparison showed that the raw 6-band data required 13000 CPU seconds, 40 man-hours, and 2580 minutes of computer connect time. The addition of two extra spectral bands (over the usual four associated with MSS data) greatly increased the time required to process the classification statistics.

The PC 3-axes data required 7000 CPU seconds, 22 man-hours and 1260 minutes of connect time (cost of data reduction technique included). This represents a 46.15 percent decrease in CPU time, a 45.00 percent decrease in man-hours and 51.16 percent decrease in computer connect time (figure 5).

Due to the technically complex and fiscally demanding nature of the project of which this study was a part, it was not possible to generate results for raw band selection. Studies performed on data collected using this same instrument, however, have indicated that raw band selection also did not improve overall classification accuracy (Gervin, et al., 1982).

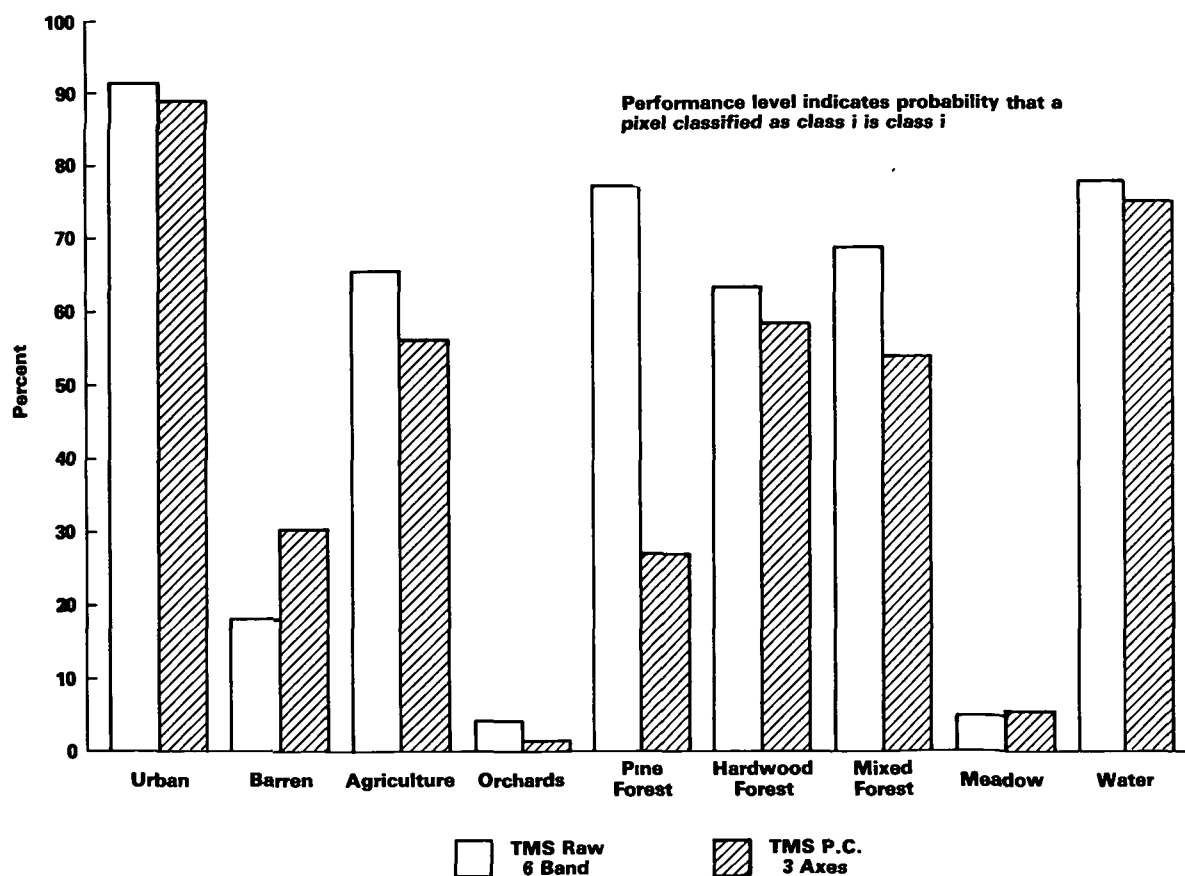
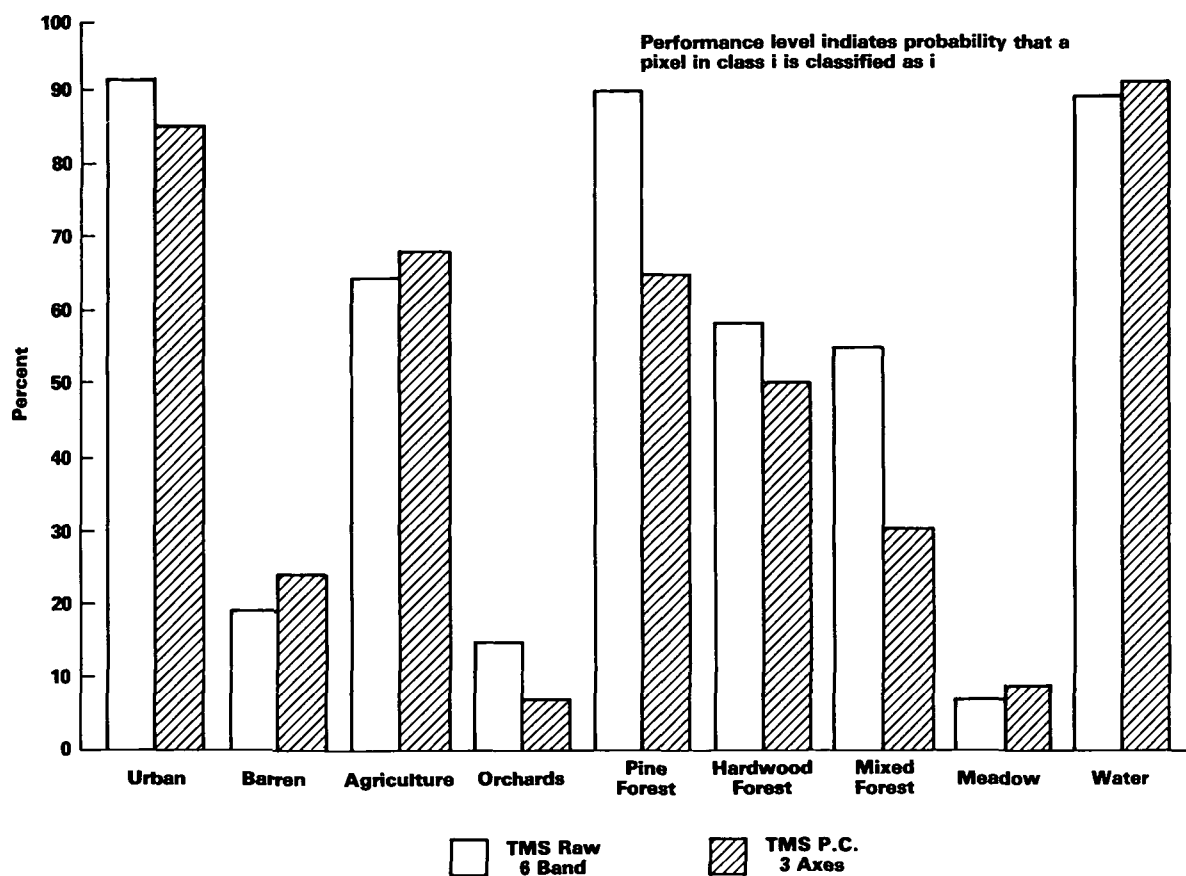


Figure 3. Comparison of TM Simulator 6 band raw and PC transformed and reduced 3 axes classification performance levels.

PROBABILITY OF CORRECTLY CLASSIFYING A PIXEL

PRINCIPAL COMPONENTS  
3 AXES

UNALTERED (RAW) 6 BAND

---

70.30%

74.61%

Figure 4. Overall performance level

COMPARISON OF WORK REQUIRED TO PROCESS  
6 BAND VS. 3 BAND (AXES) DATA

<u>6 BAND DATA</u>		<u>3 BAND (PC 3 AXES) DATA</u>	
CPU SECONDS	13,000	CPU SECONDS	7,000
MAN HOURS	40	MAN HOURS	22
CONNECT TIME (MINUTES)	2,580	CONNECT TIME (MINUTES)	1,260

SAVINGS OVER 6 CHANNEL DATA

45.15	% LESS CPU TIME
45.00	% LESS MAN HOURS
51.16	% LESS CONNECT TIME

FIGURE 5. COST COMPARISON FIGURES FOR CLASSIFYING THE RAW  
6 BAND TMS AND THE REDUCED PC TRANSFORMED DATA

## CONCLUSIONS

The principal components analysis and transformation was successful in removing noise. By concentrating the noise on lower order axes, color composite images of increased quality could be produced from axes 1, 2 and 3. In this application, data reduction effected by a principal components analysis and its transformation and the removal of the lower order axes did indeed adversely affect the overall classification accuracy. The reduction in classification accuracy, however, was minimal and may be insignificant in most applications. On the other hand, the cost savings afforded by the reduced data were substantial, > 47 percent--more than enough to offset the decrease in accuracy.

More research needs to be performed to compare raw band selection against data reduction techniques such as principal components which require transformations. It is imperative that this research be done using the actual TM data itself as recent studies appear to indicate that the TM data are quite different in character and quality from the sensors designed to simulate them. It is also important to test this procedure on a variety of target areas as the effectiveness of this and other transformations as a data reduction technique may vary depending upon the character of survey area.

## REFERENCES

1. Campbell, W. J., 1982. Integration of Remotely Sensed Data with Geographic Information Systems for Application in Energy Management. Proceedings of the 1982 Conference on Energy Resource Management, Baltimore, Maryland.
2. Gervin, J., 1982. Comparison of Land Cover Information from Landsat MSS and Airborne TMS for Input to Hydrological Models: Preliminary Results. Proceedings of the 1982 Conference on Energy Resource Management, Baltimore, Maryland.
3. Dottavio, L. L. and D. L. Williams, 1982. Mapping a Southern Pine Plantation with Satellite and Aircraft Scanner Data: A Comparison of Present and Future Landsat Sensors. Journal of Applied Photographic Engineering. Volume 8, Number 1.
4. Imhoff, M. L. and G. W. Petersen, 1980. The Role of Landsat Data Products in Soil Surveys. Institute for Research on Land and Water Resources, The Pennsylvania State University, University Park, PA.
5. Merembeck, B. and F. Y. Borden, 1978. Principal Components and Canonical Analysis for Reduction of Dimensionality of Large Data Sets. Technical Report 5-78, Office for Remote Sensing of Earth Resources, The Pennsylvania State University, University Park, PA.

OPTIMIZATION OF A NON-TRADITIONAL UNSUPERVISED  
CLASSIFICATION APPROACH FOR LAND COVER ANALYSIS

R. K. Boyd, Computer Sciences Corp.  
NASA/Goddard Space Flight Center

J. O. Brumfield  
Marshall University

W. J. Campbell  
NASA/Goddard Space Flight Center

ABSTRACT

The purpose of this paper is to analyze the conditions under which a hybrid of clustering and canonical analysis for image classification produce optimum results. The approach involves generation of classes by clustering for input to canonical analysis. The importance of the number of clusters input and the effect of other parameters of the clustering algorithm (ISOCLS) were examined. The approach derives its final result by clustering the canonically transformed data. Therefore the importance of number of clusters requested in this final stage was also examined. The effect of these variables were studied in terms of the average separability (as measured by transformed divergence) of the final clusters, the transformation matrices resulting from different numbers of input classes, and the accuracy of the final classifications.

The research was performed with Landsat MSS Data over the Hazleton/Berwick Pennsylvania area. Final classifications were compared pixel by pixel with an existing geographic information system to provide an indication of their accuracy.

The results show that both the number of clusters input to canonical analysis and the number of clusters the canonically transformed data is clustered into effect the classification accuracy. Inputting sixty clusters to canonical analysis and clustering the transformed data into thirty clusters provided the best results for the informational categories studied (urban, including commercial/industrial, and residential, agriculture, water, and surface mining) i.e., spectrally very difficult to separate classes.

A definite relationship between the number of clusters input to canonical analysis and the resulting transformation

coefficients was also observed. Specifically, those input numbers of clusters resulting in the highest level of agreement with the GIS Data also produced transformation coefficients most different from those produced by other numbers of input clusters. The separability analysis also tended to support the higher classification accuracies associated with clustering the transformed data into intermediate numbers of clusters as well as the differences associated with the number of clusters input to canonical analysis.

## INTRODUCTION

Various authors have reported significant improvements in classification accuracy associated with the use of a non-traditional unsupervised classification procedure. These accuracy improvements have been identified for both areal estimates and pixel by pixel comparisons with ground truth (Brumfield et al., 1981, Witt et al., 1982).

The procedure involves canonical analysis of the statistics derived from an iterative clustering algorithm. The transformation matrix thus developed is used to transform the original data which is then subjected to the same clustering procedures. The procedure provides all of the advantages of using clustering to derive training class statistics (and unsupervised classification in general) (Fleming and Hoffer, 1977) while at the same time incorporating the noise reduction and transformation optimization characteristics of canonical analysis (Brumfield et al, 1981).

Although the approach requires very little analyst involvement, decisions must be made regarding the number of classes input to the canonical analysis and the number of classes into which the resulting transformed data should be clustered.

The purpose of this paper is to examine the relationship between these variables and the resulting classification accuracy. Various other indicators of the performance of the procedure are also considered.

## DISCUSSION OF METHODS AND PROCEDURES

### 1. DATA SETS

The remote sensing data used in the experiments were a thirty-four (34) kilometer square subset of the Landsat MSS scene 1350-15190, dated July 8, 1973, covering the Hazleton-Berwick, Pennsylvania area. The MSS data were observed to



exhibit variable haze cover, radiometric striping, and a small amount of random noise. The study site, dissected by the Susquehanna River, is comprised of forested mountains separated by rolling valleys that have been put to a variety of agriculture usages. Major coal mining activities and the associated open pit mines, in all stages of operation, reclamation, and abandonment, are also found in the area. The industrial/commercial activities and residential sprawl of varying densities are also well represented in the area.

Two sets of color infrared photography flown in January and August of 1973 were used as reference data.

A vector (polygon formatted) data base, part of the Environmental and Land Use Data System (ELUDS) of the Pennsylvania Power and Light Company was used as ground truth for performing accuracy assessments. The following categories are coded in the vegetation/landcover layer: urban land, barren land, agricultural land, tree plantations, needle leaf forest, broad leaf forest, mixed forest, scrub land, meadow, forested wetland, unforested wetland, and waterbody.

## 2. EQUIPMENT

The experiments were carried out using the Interactive Digital Image Manipulation System (IDIMS) (Electromagnetic Systems Laboratory 1981) at the Eastern Regional Remote Sensing Applications Center (ERRSAC), NASA/Goddard Space Flight Center, Greenbelt, MD. This system consists of several components including a Hewlett-Packard Model 3000 mini-computer, a Comtal and Deanza image display terminal, a Talos coordinate digitizer table, and the associated software. The Environmental Systems Research Institute (ESRI) polygon to grid conversion software also played an important role in the research (ESRI, 1979). Canonical analysis was performed by the program CANAL developed by the Office of Remote Sensing For Earth Resources (ORSER) at the Pennsylvania State University (Turner et al, 1978).

## 3. PREPARATION OF DATA SETS

In order to allow comparison of the MSS data with the landcover information coded in the ELUDS data base the two were altered so as to correspond to a common grid system. Prior to altering the geometric characteristics of the Landsat data a histogram matching algorithm was applied to remove the six line striping in the data. The Landsat data were then resampled to a grid system referenced to the universal transverse mercator (UTM) map projection (the same map projection ELUDS polygons are referenced to). The transformation coefficients driving the resampling were derived from a third order fit of 30 ground control points

(ordered pairs of Landsat pixel addresses and UTM grid system coordinates). RMS error for these ground control points was less than 0.5 pixel. The cell size of the grid system was chosen to be 67 meters. A gridded version of the ELUDS data base, with the same UTM origin and grid cell size as the Landsat Data was created by determining for each grid cell the data value of the polygon occupying the largest part of the grid cell.

#### 4. INITIAL CLUSTERING

The first step in the procedure is to separate the remote sensing data into spectral clusters for input to the canonical analysis program.

The IDIMS program ISOCLS was used for this step. ISOCLS is a clustering algorithm which either splits or combines clusters in each iteration depending on the requirements set by the analyst for the maximum standard deviation within a cluster (STDMAX) and the minimum euclidean distance between clusters (DLMIN). ISOCLS can be seeded with class means provided by the analyst or with a single cluster defined by the mean vector of the data set to be clustered. In the latter case, this initial cluster is successively split in consecutive iterations until the resulting clusters are less variable than STDMAX. If STDMAX is set low enough the splitting will continue until the maximum number of clusters (also set by the analyst) is met; at which point ISOCLS will iterate assignment of pixels to the clusters and recalculation of the cluster mean vectors until the maximum number of iterations (set by the analyst) is reached. In this way, ISOCLS can be forced to approximate a K-means clustering algorithm (Moik, 1980).

ISOCLS was applied to the entire data set (512 lines by 512 samples). The maximum number of clusters was set to be 10, 20, 30, 40, and 60 in five separate runs. STDMAX was set at 1.5 thus forcing ISOCLS to split the initial clusters until the maximum number of clusters was reached in each case and iterate on that number of clusters as discussed above.

ISOCLS was also applied to supervised (pure) samples of water, strip mines, forest, agriculture, and urban. The supervised samples contained multiple training sites and were selected on the basis of analyst judgement to be as representative of the cover types mentioned as possible. Each sample was clustered separately, and the maximum number of clusters was set at six for each sample, resulting in 30 clusters total. This method of generating classes for input to canonical analysis is not part of the nontraditional unsupervised classification procedure and was included primarily to serve as a point of comparison.

## 5. DATA TRANSFORMATION

This part of the procedure utilizes a linear transformation of the data. The coefficients for the transformation are calculated by the canonical analysis algorithm developed at ORSER. The algorithm determines the translation, rotation, and rescaling of the data that maximizes the among cluster variability while setting the within cluster variability equal to unity (Merembeck et al., 1978). The resulting canonical transformation maximizes the separability of the clusters based upon the within cluster and among cluster variability.

The means and covariance matrices for each set of clusters derived from the procedures outlined above were input to the ORSER program CANAL to develop a transformation matrix for each set (Table II). Each transformation matrix was then input to the IDIMS program KLTRANS to perform matrix multiplications with the original data set to generate the transformed data for each case (Brumfield et al., 1981).

## 6. CLUSTERING OF TRANSFORMED DATA

The final step of the procedure is to classify the transformed data by separating the data into groups with clustering.

The transformed data sets derived from the above procedures were clustered using only the first and second transformed axes (axes one and two contain over 98 percent of the variability in the data). The STDMAX parameter in ISOCLS was set at 0.1, again forcing ISOCLS to emulate a K-means clustering algorithm. ISOCLS was used to generate 15, 20, and 30 clusters for each transformed data set discussed above. ISOCLS was also used to generate 40 clusters for the transformed data set based on 60 clusters. Table I shows the various combinations of clusters input to canonical analysis and output from clustering the transformed data sets. The clusters in each clustered transformed data set were then grouped into informational categories by comparing the cluster results with color infrared photography. Each cluster output was displayed and colored up on a color display screen to effect the comparison. The grouping process was also assisted by examination of two dimensional plots of the cluster means and covariances.

## 7. SEPARABILITY ANALYSIS

The first indicator used to check for differences related to the number of classes input to and output from the procedure was interclass separability. A modified version of the IDIMS function *diverge* was used to calculate the average transformed divergence (Swain and Davis, 1978) of those class pairs which yielded transformed divergence values less than 1500 (transformed divergence takes on values between 0 and 2000, where 2000 indicates maximum separability). This average

separability of the least separable classes was calculated for each set of clusters input to the nontraditional unsupervised procedure as well as for each set output from the procedure and is graphed in Figure I.

#### 8. DETERMINATION OF LEVELS OF AGREEMENT

The second indicator of differences associated with the number of classes input to and output from the procedure was level of agreement with ground truth. The land cover layer of the ELUDS Data Base served as ground truth for this study.

The classes in each clustered transformed data set were grouped into five informational categories (urban, strip mines, agriculture, forest, and water) for comparison with the ELUDS landcover information. The grouping was accomplished by renumbering each cluster in each clustered transformed data set to the number chosen to represent the assigned category. The 12 ELUDS landcover classes were grouped into the same informational categories and renumbered to reflect the same coding scheme. Each renumbered clustered transformed data set was then compared pixel by grid cell with the renumbered ELUDS landcover layer to produce a contingency table showing the number of pixels in agreement and disagreement by category. Percentages of agreement were calculated by category and are shown in Table III. Percentage of agreement was calculated by dividing the number of pixels in agreement for the category in question by the total number occurring in the data base for that category. Overall agreement was calculated by dividing the total number of pixels correctly classified by the total number in the data base. These figures are being referred to as levels of agreement instead of accuracy because of the fact that ground verified test sites were not used to calculate them. The ELUDS Data Base is undoubtedly fairly accurate. However, to the knowledge of the authors, no quantitative estimate of its accuracy exists.

### RESULTS

#### 1. TRANSFORMATION COEFFICIENTS

The transformation coefficients for Axis 1 and Axis 2 resulting from canonical analysis of the various numbers of input classes are shown in Table II. The coefficients seem to fall into four unique sets, those based on 10 clusters, those based on 20, 30 and 40 clusters, those based on 60 clusters, and those based on the 30 clusters from supervised samples. Without question both the number and source of the class statistics input to canonical analysis affect the resulting transformation coefficients.

## 2. SEPARABILITY ANALYSIS

The average separability of the least separable classes for each set of input and output clusters is shown in Figure I. Three main trends can be seen from this graph. First, the average separability of the least separable classes tends to increase as the number of clusters increases. Second, for any given number of output classes the average separability of the output classes is constant or decreases slightly as the number of input classes increases. Third, there is a slight increase in separability as the number of output classes is increased for any given number of input classes. Unfortunately, the magnitude of the third trend cannot be viewed as being significant due to the inherent variability associated with calculating transformed divergence from class statistics (Swain and King, 1973). Interestingly, this increased separability of the 60 cluster set of input classes over the 20, 30 and 40 cluster sets (first trend) does concur with the changes observed in the transformation coefficients resulting from those sets.

## 3. PERCENTAGE OF AGREEMENT

The percentage of agreement of each set of output classes with the ELUDS Data Base is shown in Table III. As is evidenced by the low percentages of agreement for urban and barren, separating these categories from the other categories with MSS Data in this area is very difficult. However, of greater relevance to the scope of this paper are the trends observed in the levels of agreement. Perhaps the most obvious difference is the difference in overall agreement between 15 classes output and 20 or 30 classes output. This decreased overall agreement is consistent with the decreased average separability discussed earlier. The three highest overall levels of agreement were obtained from the 60/30, 60/40, and 30(supervised)/30 sets. Furthermore, with the exception of the 10/20 set, the highest agreement for barren were also obtained with the 60/30, 60/40, and 30(supervised)/30 sets. The results also show that there is an interplay between the number of classes input and the number of classes output. Although, 10 classes input produced an overall level of agreement of 74.7 percent for 20 classes output, it produced only 72.7 percent for 15 classes and 30 classes output. Similarly 60 classes input produced 16.4 and 12.5 percent agreement for barren for 30 and 40 classes output but 0 percent for 20 classes output. Finally, the source of the classes has a definite influence on level of agreement. Thirty input classes from supervised samples produced higher overall agreement than 30 input classes from a systematic sample of the data, regardless of the number of output clusters used in the latter case.

## CONCLUSIONS

Clearly the number and source of classes input to and output from the nontraditional unsupervised technique has an impact on the resulting classification accuracy. The results indicate the best overall classification will be obtained when the classes input to canonical analysis sufficiently subdivide the total spectral variability in the data set. In this experiment it was necessary to cluster a systematic sample of the data into 60 clusters or separately cluster supervised samples of the data into six clusters each to accomplish that subdivision. Although certain lower numbers of input classes may produce good results when used in combination with certain other numbers of output classes (e.g. 10/20 in this experiment) it will be difficult to predict these combinations in advance. By subdividing the data set into a large number of clusters the likelihood of representing spectral groupings associated with informational categories is increased. The results also show that separating the transformed data into an intermediate number of clusters is sufficient to obtain the best classification. In this experiment no significant increase in level of agreement was obtained as the number of output classes was increased from 30 to 40. Furthermore, comparable results were obtained when 30 classes were output from the transformed data based on 60 clusters from a systematic sample and from the transformed data based on 30 clusters from supervised samples.

The optimum numbers of clusters will undoubtedly vary from data set to data set. However, it is doubtful that any data set will contain categories more difficult to separate than urban, strip mines, agriculture, and water as contained in the data set used in this experiment. On this basis the 60/30 combination should provide nearly optimal results for any MSS data set.

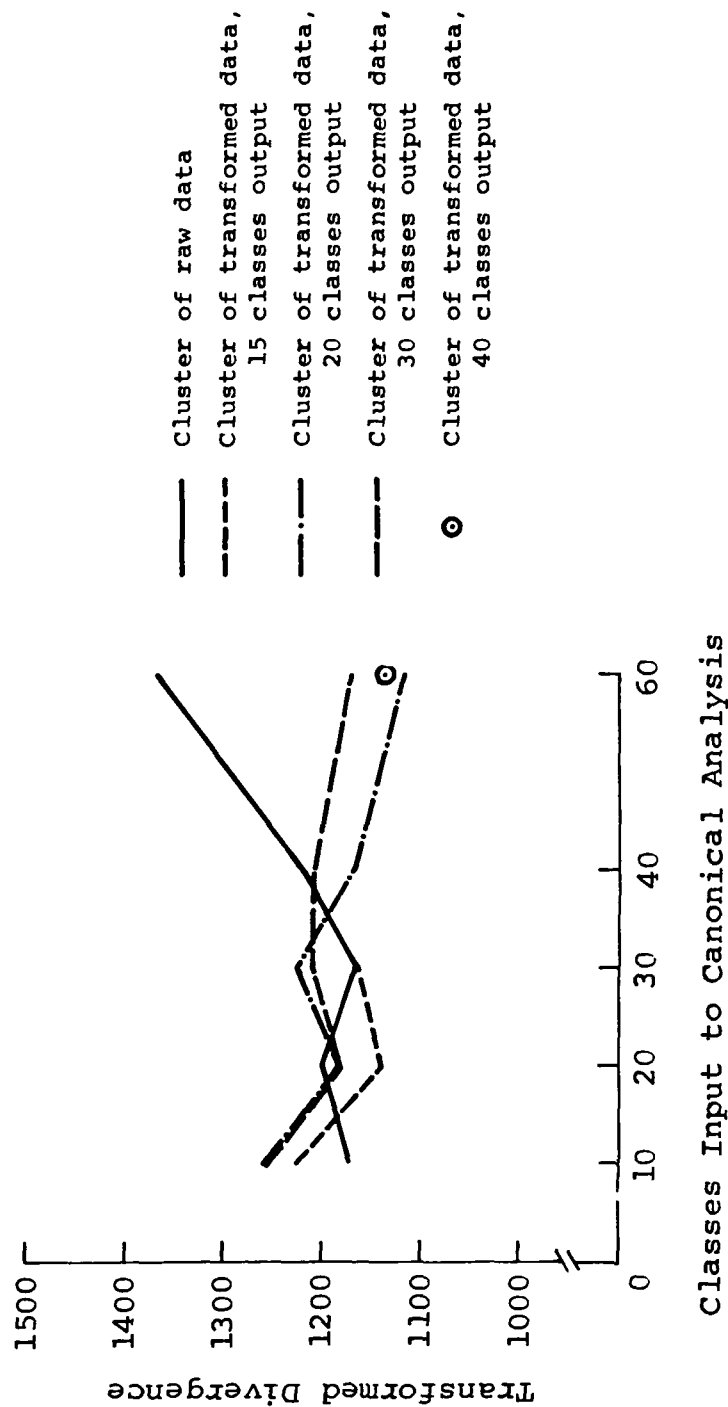


FIGURE I. Average separability of the least separable classes for each set of clusters input to and output from the nontraditional unsupervised classification approach.

TABLE I.

Combinations of number of clusters  
input to/ and output from the  
nontraditional classification procedure.

		Input Clusters				
O u t p u t  C l u s t e r s		10	20	30	40	60
	15	X	X	X		
	20	X	X	X	X	X
	30	X	X	X,S	X	X
	40					X

X-Input clusters generated by clustering entire data set.  
S-Input clusters generated by clustering supervised samples  
of the data.



TABLE II.

Transformation Coefficients  
for axes one and two, as  
produced by canonical analysis.

		Axis 1		Band 3	Band 4
		Band 1	Band 2		
I n p u t	10	-.1193	-.0260	.0774	.0670
	20	-.2471	-.1096	.1189	.2525
	30	-.2496	-.1686	.0743	.3607
	40	-.2405	-.1639	.0644	.4217
	60	-.1183	-.1610	.2535	.4760
	30 <sup>1</sup>	-.0988	-.0344	.1775	.2820
C l u s t e r s		Axis 2		Band 3	Band 4
		Band 1	Band 2		
	10	-.0320	.1386	.0769	-.0128
	20	.0265	.2879	.1752	-.0142
	30	.0540	.3468	.1794	.0097
	40	.0788	.3538	.2319	-.0182
	60	.2451	.4197	.2394	-.0086
	30 <sup>1</sup>	.2096	.2146	.1148	-.0251

<sup>1</sup>From supervised samples.

TABLE III.

Percentage of agreement of  
classifications with the  
ELUDS data base.

Input/ Output	Urban	Strip	Agri.	Forest	Water	Overall
10/15	19.8	0.0	57.2	93.3	73.1	72.7
20/15	20.8	0.0	57.4	92.2	72.6	72.1
30/15	22.5	0.0	57.4	93.3	72.7	72.9
10/20	22.3	20.9	75.5	88.2	53.9	74.7
20/20	16.3	0.0	77.5	89.3	70.3	74.5
30/20	16.0	0.0	75.9	90.5	70.5	74.9
40/20	16.0	0.0	77.2	89.4	70.2	74.5
60/20	13.5	0.0	79.1	89.4	67.4	74.7
10/30	18.7	0.8	68.4	89.4	66.8	72.7
20/30	31.3	3.4	60.6	93.5	65.4	74.3
30/30	31.2	3.1	64.1	92.0	66.4	74.2
40/30	30.1	3.5	65.8	91.0	65.1	73.9
60/30	27.1	16.4	72.3	90.1	63.5	75.5
60/40	24.2	12.5	70.7	91.5	65.8	76.1
30 <sup>1</sup> /30	25.4	15.8	73.4	90.9	63.8	75.9

<sup>1</sup>Generated by clustering of supervised samples.

## REFERENCES

1. Brumfield, J. O., H. L. Bloemer, W. J. Campbell, "An Unsupervised Approach for Analysis of Landsat Data to Monitor Land Reclamation in Belmont County, Ohio", Seventh International Symposium for Machine Processing of Remotely Sensed Data.
2. Electromagnetic Systems Laboratory, "IDIMS Functional Guide", 1981.
3. Environmental Systems Research Institute, "Environmental and Land Use Data System (ELUDS)," 1978.
4. Environmental Systems Research Institute, "Grips User Manual", 1979.
5. Fleming, M. D. and R. M. Hoffer, "Computer Aided Analysis Techniques for an Operational System to Map Forest Lands Utilizing Landsat MSS Data." Laboratory For Applications of Remote Sensing, 1977.
6. Merembeck, F., F. Y. Borden, M. H. Podwysocki, and D. N. Applegate, "Application of Canonical Analysis to Multispectral Scanner Data." 14th Annual Symposium on the Application of Computers in Mine and Industry, 1977.
7. Moik, J. G., "Digital Processing of Remotely Sensed Images." National Aeronautics and Space Administration, 1980.
8. Turner, B. et al., "Satellite and Aircraft Multispectral Scanner Digital Data User Manual," Office for Remote Sensing of Earth Resources, 1978.
9. Swain, P. H., and Davis, S. M., "Remote Sensing The Quantitative Approach." McGraw-Hill International Book Company, 1978.
10. Swain, P. H., and King, R. C., "Two Effective Feature Selection Criteria For Multispectral Remote Sensing." Laboratory for Applications of Remote Sensing, 1973.
11. Witt, R. G., H.H.L. Bloemer, Beldeon Bly, J. O. Brumfield, W. J. Campbell, "Comparing Digital Data Processing Techniques for Surface Mine & Reclamation Monitoring", American Society of Photogrammetry annual meeting, April, 1982.

# USING HCMM THERMAL DATA TO IMPROVE CLASSIFICATION OF MSS DATA

R. G. WITT  
NASA/Goddard Space Flight Center

R. S. SEKHON  
Computer Sciences Corporation/  
Goddard Space Flight Center

Greenbelt, Maryland 20711, U.S.A.

## ABSTRACT

Spectral overlap between urban and rural land use/land cover categories can lead to unacceptable map accuracy levels in the classification of Landsat multispectral scanner (MSS) data. The four MSS bands used alone are not always adequate to distinguish among various land uses and cover types having similar spectral responses. This study investigates the use of thermal data from the Heat Capacity Mapping Mission (HCMM) satellite as a means of improving MSS land cover classification accuracies for urban versus rural categories.

## 1. INTRODUCTION

The Heat Capacity Mapping Mission (HCMM) satellite was launched on April 26, 1978 to acquire thermal and reflectance data for potential applications in a variety of disciplines. A number of investigators have reported on the utility of the data for discriminating geologic types (Cole and Edmiston, 1980), mapping soil moisture (Reginato et al., 1976; Kocin, 1979), measuring plant canopy temperatures (Harlan et al., 1981; Wiegand et al., 1981), and studying patterns of thermal circulation in large water bodies (Schowengerdt, 1982). While the potential for using satellite-acquired thermal data to detect and study urban heat islands has been explored by Carlson et al. (1977), Matson et al. (1978), Price (1979), and Rao (1972), there has been no practical application of HCMM data to delineate urban areas using digital classification techniques. This paper documents classification procedures for using HCMM data along with Landsat MSS data in order to improve the separability of urban and non-urban areas.

## 2. DESCRIPTION OF THE HCMM SATELLITE AND DATA

The HCMM satellite, whose mission lasted until October 1980, carried a two-channel radiometer to sense emitted data

in a thermal infrared (IR) band (10.5-12.5  $\mu\text{m}$ ) and reflected data in a visible band (.5-1.1  $\mu\text{m}$ ). The thermal channel's NEAT is 0.3°K at 280°K, with a nominal spatial resolution of 600 meters at nadir for both bands. The three types of data obtained by HCMM were reflectance and infrared data during a daytime pass, and infrared data at night. Thermal inertia and temperature difference (day minus night) data were also calculated. These data sets are available individually or as registered day/night pairs in image format and on computer-compatible tapes (CCTs).

The HCMM satellite thermal sensor was calibrated at launch to measure a range of temperature values between 260°K and 340 K (-13°C to 67°C, or 8.6°F to 152.6°F). With an eight-bit (0-255) configuration, the sensitivity of the thermal channel is such that HCMM was capable of measuring .3°K or less than .6°F changes in temperature. This high thermal sensitivity suggests that it should be possible to differentiate between relatively dense man-made materials and surfaces, and natural cover types, vegetation and water, on the basis of their relative emissivities.

### 3. RATIONALE FOR COMBINED CLASSIFICATION

This study was an extension of a land use/land cover change detection project conducted by the Eastern Regional Remote Sensing Applications Center (ERRSAC) of NASA/Goddard Space Flight Center, and the Ohio Environmental Planning Agency/Office of the Planning Coordinator (EPA/OPC). Clark County, a largely agricultural county in west-central Ohio, was the area selected for classification of two MSS data sets. The city of Springfield (population 70,000) and two other large towns are the only extensive urban developments within Clark County. When the initial classifications of MSS data, both unsupervised and supervised, failed to distinguish between commercial-industrial areas and bare agricultural lands, as well as lower density residential areas and cropland, a decision was made to integrate HCMM thermal data. The underlying assumption made was that the emitted temperatures associated with urban land uses are generally higher than those of surrounding rural land covers, and that this dichotomy would allow urban areas to be delineated on the basis of characteristic temperature differences.

The data sources used included the following:

- Landsat MSS scenes imaged on 2 June 1974 and 26 May 1977.
- HCMM digital data sets (daytime thermal IR and visible) acquired 12 March 1979.
- Ohio Department of Natural Resources (ODNR), Ohio Capability Analysis Program (OCAP) 1:24,000 scale level I land use map of Clark County.

- U. S. Geological Survey 7.5' topographic quadrangle maps of Clark County.
- Aerial photography at 1:60,000, 1:30,000, and 1:24,000 scales for portions of Clark County.

The OCAP level I (Anderson et al., 1976) land use map represented the existing land use classification encoded in the State's data base, and was used as the main source of ground truth information in carrying out the accuracy assessment.

#### 4. PROCEDURES

Three classifications were developed in order to compare how the addition of HCMM data influenced the map accuracy of the results. A classification of the two MSS scenes covering Clark County was the first stage of the study. Unsupervised signatures were developed by using a clustering technique, and supervised signatures were derived by selecting training sites. The final signatures were input to a maximum likelihood classifier to produce classifications of the entire County. When the results did not prove satisfactory, it was found that changing or eliminating certain ambiguous signatures did not lead to significant improvement.

Two general approaches, referred to herein as the masking and merging techniques, were used to integrate the HCMM data. With the masking technique, HCMM thermal data were density sliced to create binary masks representing urban/non-urban areas of the image. The 'urban' and 'rural' masks then were multiplied by the raw MSS data set to create two complementary images. These were classified using separate sets of signatures, and the resulting images were recombined to form the final classification. In the merging technique, two HCMM bands (day infrared and visible) were combined with the four MSS bands, and a subset of the merged data set was input for unsupervised clustering. The resulting 64 clusters were labelled, and the same maximum likelihood algorithm was applied to the entire image in the final classification. Throughout this paper, the three techniques are referred to as the MSS only, MSS-HCMM masked, and the MSS-HCMM merged classifications.

##### 4.1. MSS Only Classification

Subsections of the geocorrected Landsat data sets for the Clark County study area were sent to the Ohio EPA/OPC. The agency analyzed the digital Landsat data by using the ORSER-OCCULT software package. The OCCULT software is a user-friendly and conversational system for utilizing the ORSER software and its modifications. The ORSER-OCCULT package includes a series of analytical routines which are used to develop signatures (sets of means and standard deviations)

from known, uniform areas in the subscene, and classifies the entire study area based on these signatures.

Classification results were modified by using progressively refined signatures to achieve a classification comparable to the ground truth information. The training site statistics developed by the Ohio EPA/OPC were sent to ERRSAC for classification there with the Interactive Digital Image Manipulation System (IDIMS) software on a Hewlett-Packard (HP) 3000 computer. Using these signatures, the entire subscene was classified with a maximum likelihood algorithm to duplicate the classification at ERRSAC. All spectral classes were renumbered to the six land use/land cover categories and reclassified to smooth out single-pixel discrepancies.

Visual inspection of this classified image found unacceptable rates of confusion between urban areas and bare fields that were being classified mistakenly as either commercial-industrial or residential land use. The first means of correcting this problem was to attempt a minimum distance classification. Critical limits were raised for agriculture/bare soils signatures and, when this did not result in significant improvement, limits for commercial-industrial and residential signatures were lowered. These changes had little effect on the classification. All signatures were then evaluated individually to detect and discard those which were causing the most confusion. The resulting classification still showed no major improvement, and the MSS only classified image did not undergo any further refinement.

#### 4.2. MSS-HCMM Masked Classification

Data from the HCMM satellite, scene ID AA0320-18200 (image center at N39 55' and W82 15'; sun elevation of 45 ), were received from the National Space Science Data Center at Goddard Space Flight Center. Both data sets acquired, day thermal infrared (day IR; 10.5-12.5 m) and day visible (day VIS; 0.5-1.1 m), were taken simultaneously at about 1:30 in the afternoon, the time of maximum surface temperature. The image quality was good, and though there was ten per cent cloud cover, no clouds or haze obscured the study area. The digital images came registered to one another, but not to any map base. Thus the images had to be registered to available maps of the study area, and resampled using a nearest neighbor procedure to overlay the MSS data at the same scale and pixel size (50 meters). Appropriate subsets corresponding to the MSS image of the study area were created. After doing a contrast stretch of both the infrared and visible bands to the fullest range possible, it was determined that the thermal band offered better definition of urban areas when compared with maps of Clark County. Therefore, the thermal image was selected to be used in developing the binary urban and rural masks.

The binary image was created by mapping the thermal IR data set to desired levels. A stochastic method was employed to determine the threshold which represented the boundary between the higher emissivity of warmer, urban areas, and the lower emissivity of cooler, non-urban areas. All digital values above this level were mapped (that is, renumbered) to one, and all digital values at this level and below were mapped to zero. After repeated experimentation with the mapping, the values 46 and below were mapped to zero and the values 47 and above were mapped to one, creating an urban area mask. The rural area mask was complementary to the latter image; it was obtained by mapping 0 1 values to 1 0 values. These masks were used to create complementary MSS images representing urban and non-urban areas. Each of the four MSS bands was multiplied by the urban mask and then united to form a masked MSS image representing only urban areas. Similarly utilizing the MSS image and the non-urban mask, a masked MSS image representing rural areas was formed.

Separate groups of signatures were used to classify the 'urban' and 'rural' data sets. The urban group of signatures included commercial-industrial, residential, agriculture, forest and water signatures (all except bare soil), while the rural group included agriculture, bare soil, forest and water. No pixels in the rural image could be classified as commercial-industrial or residential, thus eliminating some small towns and low-density suburban areas. However, urban areas did encompass some farmland and parkland, which were classified correctly as agriculture in the broad sense.

After the 'urban' and 'rural' images were classified separately using a minimum distance algorithm with modified threshold limits for all signatures, three additional steps were necessary to complete the classification. Each of the classified images was renumbered to reflect the final number of land use/land cover categories, the complementary images were added together, and the united classification was reclassified to smooth out single-pixel discrepancies.

#### 4.3. MSS-HCMM Merged Classification

The third classification was generated from a Landsat MSS-HCMM merged data set. The registered and resampled HCMM subscene was united with the MSS subscene to form a six-band image, containing four MSS bands and two HCMM bands (day IR and visible). Unsupervised signatures were developed from a representative test area using a clustering technique. These signatures then were applied to the entire study area using a maximum likelihood classifier. The spectral classes were labelled as the same six land use/land cover categories in the other two classifications. The classified image was renumbered and reclassified as previously described, thus completing the processing phase of the study.



#### 4.4. Accuracy Assessment Procedures

The accuracy assessment was conducted by comparing land cover maps from the three classifications with the 1:24,000 scale OCAP land use map of level I categories. Three U.S.G.S. quadrangles in central Clark County - Springfield, New Moorefield, and Donnelsville - were selected for this procedure. Thematic grayscale overlays of each classification were output on a Versatec plotter at the same scale to facilitate the comparison.

Each of the thematic overlays (one per quad for three classifications, or a total of nine) was registered to the OCAP land use map on a light table. A grid with a cell size of 25 pixels (5x5 pixel blocks) was superimposed on the maps, and numbered along its x and y axes. Using a random numbers table, 50 cells were sampled on each overlay and compared with the ground truth map on a pixel-by-pixel basis. A count of the number of correct and incorrect pixels was kept; these results are presented in Tables I and II.

Because the Landsat and OCAP categories were not identical, some categories were combined for the accuracy assessment. The Landsat commercial-industrial and residential categories were combined so that they could be compared with urban/built-up and barren land categories in OCAP. The agriculture and bare soil categories (Landsat) also were added together and compared with OCAP's combined agriculture and rangeland. Forest and wetlands were combined (OCAP) and compared with forest in Landsat. The remaining level I category was water; thus it was possible to compare four general land cover types.

#### 5. RESULTS

The results of the three classification procedures are presented below in two tables. Table I compares the classification acreage counts for the four level I categories with the ground truth acreages from OCAP, for the total area of the three 7.5' quads on which the accuracy assessment was conducted. Table II shows the percentage of agreement derived from the accuracy assessment, again for all categories and the three quads totalled.

From the first table it can be seen that the MSS-HCMM merged classification comes closest to estimating the actual extent of urban and agricultural lands, according to the OCAP information. Only the forest category is overestimated, and this may be a result of cluster mislabelling. It became evident during the accuracy assessment that many pixels classified as forest on the HCMM merged classification should have been agriculture or urban instead. If this problem were to be corrected, the agreement between the OCAP and HCMM merged acreages would be even greater.

Table I: ACREAGE COUNTS BY CLASSIFICATION

Level I Category	MSS only (1977)	HCMM masked (1977-79)	HCMM merged (1977-79)	OCAP (1979)
Urban	30706 28%	11958 10.9%	17985 16.4%	19834 18.1%
Agric.	67101 61.1%	89043 81.1%	75278 68.6%	78524 71.5%
Forest	7430 6.8%	6804 6.2%	14545 13.2%	8482 7.7%
Water	2026 1.8%	1997 1.8%	1993 1.8%	2942 2.7%
Unclas.	2537 2.3%	7 -	- -	- -
Totals	109800	109802	109801	109782

Conversely, in terms of locational accuracy, the HCMM merged classification delineated forest better than the other techniques (56.3 per cent, as opposed to 52.9 per cent and 37.7 per cent for the MSS only and HCMM masked classifications). This reflects a low rate of omission errors and a high rate of commission errors for the forest category, which led to decreased locational accuracies for the urban and agricultural categories in the MSS-HCMM merged classification. In general, however, the MSS-HCMM merged classification had the virtues of the other two classifications without the faults of either. It provided more highly correlated estimates of urban and agricultural lands, and locational accuracies as good as or better than the MSS only classification.

The other fact that emerges from the comparison of acreage counts is that the HCMM masked classification underestimates urban land and overestimates agricultural land by approximately the same amounts that the MSS only classification does the reverse. For urban land, the HCMM masked estimate is 7.2 per cent low and the MSS only estimate is 9.9 per cent high; while for agriculture the MSS only estimate is 10.4 per cent low and the HCMM masked estimate is 9.6 per cent high. This is the result of a fairly indiscriminant urban classification with the MSS data (high rates of bare soil being classified as commercial-industrial, and cropped fields being mapped as residential areas), and a very restrictive urban classification on the HCMM masked image. There were very few

errors of commission in the urban category for the HCMM masked classification as a result (see Witt and Sekhon, 1982). The fact that only areas within the urban mask could be classified as commercial-industrial or residential led to the relatively high locational accuracies for urban (65.7 per cent) and agriculture (87.9 per cent) on the HCMM masked classification.

Table II: ACCURACY ASSESSMENT RESULTS

Level I Category	Classification Percentage Correct*		
	MSS only	HCMM merged	HCMM masked
Urban	61.8	60.0	65.7
Agric.	82.0	84.0	87.9
Forest	52.9	56.3	37.7
Water	81.4	89.1	81.6
Totals	73.8	76.1	79.6

(\*calculated as 100% minus the average of commission and omission errors)

Thus, the MSS-HCMM masked classification had the highest overall per-pixel accuracy for the three 7.5' quadrangles checked, while the MSS-HCMM merged classification better represented overall acreage totals for the two principal categories assessed. Furthermore, the MSS-HCMM merged classification was clearly superior in terms of locational accuracy for the forest and water categories, while the MSS only classification included more than 2 per cent unclassified data.

## 6. CONCLUSIONS

The addition of HCMM thermal day infrared (and visible) data to the MSS classification did lead to better results in map accuracy for level I land use/land cover categories. Although neither the masking nor the merging procedure led to dramatic increases in classification accuracy, both techniques show potential for improved delineation of urban land from surrounding non-urbanized areas.

The masking technique was more effective in delineating the city of Springfield and larger towns, while excluding small towns and linear developments which were below the spatial resolution of the HCMM sensor. Thus, small communities which were not "bright" or warm enough to saturate a single/multiple HCMM pixels could not be classified as urban using the binary masking (urban/rural stratification) technique.

On the other hand, the merging technique relied on the thermal information that was added directly to the classification of every MSS pixel. Obviously, large blocks within the MSS data set had the same thermal value due to the coarser resolution of HCMM, but variation of the four MSS bands was enough to eliminate any appearance of blockiness in the final classified image. The clustering method by which the merged classification was executed allowed smaller towns and major transportation corridors to be properly labelled as urban even if they were below the HCMM spatial resolution. There was still some confusion, however, between urban and rural land cover categories associated with the cluster labelling process, which resulted in locational accuracies lower than those anticipated by the researchers.

It may be possible to further improve classification results by using a hybrid procedure incorporating both techniques. In this procedure, a binary (urban/non-urban) image would be created first from the HCMM thermal data. After multiplying the binary image with the merged MSS-HCMM data set, the resulting 'urban' and 'rural' six-banded images would be subjected to separate clustering and cluster labelling. The classification process then would reflect a bias toward urban land uses within the 'urban' image, and non-urban land covers within the 'rural' image.

## 6.2. Additional Research

Future research relating to the integration of HCMM and MSS data for improved classification results should focus on several key topics. More work needs to be done to determine optimal conditions under which to employ HCMM data for urban area delineation. This probably is dependent not only on the density and extent of the particular urban area, but on such factors as the time of year of the HCMM and MSS images, atmospheric conditions on a given day, and the types of vegetation present in the surrounding area. Some combination of these factors may determine whether it is appropriate to utilize HCMM data for delineating various types of urban areas, or whether (for example) the analysis would profit more from the digitization of urban area boundaries.

Research is now being carried out to test the radiometric stability of HCMM data for the same scene from date to date. Because of the problems with the absolute calibration of the HCMM thermal sensor, the non-experimental usage of HCMM thermal data has been subject to question. If it can be proven that there is little variation in sensor performance over time, it is expected that the use of HCMM thermal data for the type of application discussed above might become more widely accepted.

## REFERENCES

1. Anderson, J. R., Hardy, E. E., Roach, J. T., and R. E. Witmer. 1976. A land cover classification system for use with remote sensing data. U. S. Geological Survey Professional Paper 964, Washington, D.C., 33 p.
2. Augustine, J. A. 1978. A detailed analysis of urban ground temperature and albedo using high-resolution satellite measurements. M.S. thesis, Department of Meteorology, Pennsylvania State University (DM/PSU).
3. Boland, F. E. 1977. A model for determining surface temperatures and sensible heat fluxes over the urban-rural complex. M.S. thesis, DM/PSU.
4. Carlson, Toby N. 1980. Applications of HCMM satellite data to the study of urban heating patterns: remote estimate of the surface energy flux, moisture availability and thermal inertia over urban and rural terrain. NAS5-24264 Final Report, DM/PSU, College Park, Pa., 62 p.
5. \_\_\_\_\_, Boland, F. E., and J. A. Augustine. 1977. Potential application of satellite temperature measurements in the analysis of urban land use over urban areas. Bulletin of the American Meteorological Society, 58:1301-1303.
6. Cole, M. M., and D. J. Edmiston. 1980. HCMM and Landsat imagery for geologic mapping in Northwest Queensland, in Fourteenth International Symposium on Remote Sensing of Environment, San Jose, Costa Rica, April 23-30, 1980, Environmental Research Institute of Michigan, Ann Arbor, Mich., Proceedings: Vol. 3, pp. 1849-1857.
7. DiCristofaro, D. C. 1980. Remote estimation of the surface characteristics and energy budget over an urban-rural area and the effects of surface heat flux on plume spread and concentration. M.S. thesis, DM/PSU, 106 p.
8. Dodd, J. K. 1979. Determination of surface characteristics and energy budget over an urban-rural area using satellite data and a boundary layer model. M.S. thesis, DM/PSU, 87 p.
9. Harlan, J. C., Rosenthal, W. D., and B. J. Blanchard. 1981. Dryland pasture and crop conditions as seen by HCMM. NAS5-24383 Final Report 3712, Texas A&M University Remote Sensing Center, College Station, Texas, 53 p.
10. Kocin, P. J. 1979. Remote estimation of surface moisture over a watershed. M. S. thesis, DM/PSU, 62 p.

11. Matson, Michael, McClain, E. P., McGinniss, D. F., and J. A. Pritchard. 1978. Satellite detection of urban heat islands. *Monthly Weather Review* 106:1725-1734.
12. NASA/Goddard Space Flight Center. 1978 (revised 1979, 1980). Heat Capacity Mapping Mission Data Users' Handbook for Applications Explorer Mission-A (AEM). 120 p.
13. Price, J. C. 1979. Assessment of the urban heat island effect through the use of satellite data. *Monthly Weather Review* 107:1554-1557.
14. Rao, P. K. 1972. Remote sensing of urban heat islands from an environmental satellite. *Bulletin of the American Meteorological Society* 53:647-648.
15. Reginato, R. J., Idso, S. B., Vedder, J. F., Jackson, R. D., Blanchard, M. B., and R. Goettelman. 1976. Soil water content and evaporation determined by thermal parameters obtained from ground-based and remote measurements. *Journal of Geophysical Research* 81:1617-1620.
16. Schowengerdt, Robert. 1982. Enhanced thermal mapping with Landsat and HCMM digital data, in American Congress on Surveying and Mapping-American Society of Photogrammetry 48th Annual Meeting, Denver, Colorado, March 14-20, 1982, Proceedings: American Society of Photogrammetry, Falls Church, Va., pp. 414-422.
17. Short, N. M. 1981. The Heat Capacity Mapping Mission, in Second Eastern Regional Remote Sensing Applications Conference, Danvers, Mass., March 9-11, 1981, Proceedings: NASA Conference Publication 2198, Goddard Space Flight Center, Greenbelt, Md., pp. 1-5.
18. \_\_\_\_\_, and L. M. Stewart, Jr. 1982. The HCMM Anthology. NASA publication in press.
19. Wiegand, C. L. 1981. Plant cover, soil, temperature, freeze, water stress, and evapotranspiration conditions. AgRISTARS EW/CCA Final Report EW-UI-04103, JSC-17143, 159 p.
20. Witt, R. G., Sekhon, R. S., and Anthony Sasson. 1982. An urban-rural land use/land cover inventory of Clark County, Ohio, using Landsat digital data. Eastern Regional Remote Sensing Applications Center (ERRSAC) open file project report, NASA/Goddard Space Flight Center, Greenbelt, Md., 30 p.

THE USE OF PRINCIPAL COMPONENTS FOR  
FOR CREATING IMPROVED IMAGERY FOR  
GEOMETRIC CONTROL POINT SELECTION

MARC L. IMHOFF  
NASA/Goddard Space Flight Center  
Greenbelt, Maryland 20771, U.S.A.

ABSTRACT

A directed principal component (PC) analysis and its transformation was applied to 7-channel Thematic Mapper Simulator (TMS) data and 4-channel Landsat multispectral scanner system (MSS) data collected over the city of Lancaster, Pennsylvania, to create improved imagery for geometric control point selection for image to image registration.

The analysis was controlled so that the transformation matrix was generated from statistics gathered only on the urban and high density residential areas in order to enhance the infrastructural features desired for geometric control point selection.

Nineteen temporally stable geometric control points, such as road intersections and bridges, were selected for a 236 km<sup>2</sup> area using USGS 7½ minute topographic quadrangles and color infrared photography. The control points were visible on both the TMS and MSS imagery. On the first attempt the corresponding image control points were selected on both data sets without using the principal components transformation. Many of the road intersection locations were visible but the actual road crossings could not be distinguished. As a result, mensuration errors using raw data exceeded the equivalent of two (79 x 79 m) pixels. The application of a guided principal components transformation yielded TMS and MSS single band images showing improved detail in the scene's urban and residential infrastructure. The PC transformed data sets were then utilized for the reselection of geometric control points. By showing greater detail, control points on both the TMS and MSS imagery could be located with greater precision using the PC transformed data. Control point reselection after transformation resulted in a 50 percent decrease in registration error.

1.0 INTRODUCTION

Accurate geometric correction and cartographic registration is an important factor in making use of remotely sensed image data. In an era of increasing interest and development of geographic information systems (GIS) it is imperative that the many data layers included in these systems be geometrically matched or registered to one another. In the case of remotely sensed imagery this becomes particularly important not only for ultimate input to a GIS but also for the creation of merged multisource data sets suitable for spectral and spatial analysis.

Image to image geometric registration can be complicated by a number of basic factors:

1. spatial differences due to differing ground instantaneous fields of view (GIFOV) between data collection instruments,
2. platform differences (satellite vs. non-satellite, etc.),
3. spectral differences (i.e., differing number of bands and band widths).

When dealing with data of two or more differing spatial resolutions, one is confronted with the problem of multiple resampling. A resampling must occur in order to produce the images in a common pixel size and then again to geometrically register the images to a particular geometric base. Even when that process is accomplished in one routine, the data is degraded. Various resampling techniques represent tradeoffs in their effects on image geometry as well as the density numbers (DN) associated with each pixel (T. L. Logan and A. H. Strahler, 1979). Nearest neighbor best preserves original DN values but creates a moiré pattern in the imagery causing registration accuracy to vary considerably throughout the image (Jayroe, 1976). Cubic convolution and bilinear interpolation appear to be the best techniques for resampling imagery for input to a GIS; however, in both cases the original (DN) values of the data are altered. A comprehensive review of interpolation techniques, complete with an excellent list of references can be found in a work by Billingsley (1982).

Another problem caused by differing spatial resolutions between imagery is areal measurement. Even after resampling an image the basic areal extent of a particular target feature will remain the same as it was before resampling. A river system or roadway, for example, imaged by two instruments with different IFOV's will record two different areal measurements associated with those same features. This becomes a particularly annoying problem when one overlays imagery containing such features as winding rivers and/or roadways where not only do their widths vary but the concave-convex curve areas for those targets differ appreciably as well.

Platform differences can also cause complications. Imagery collected from an airborne instrument platform will have a number of geometric and radiometric distortions due to sun angle, aircraft pitch, yaw, roll, aircraft and instrument jitter, etc., that are only minimally present in satellite data. Unfortunately, these effects are of fairly high spatial frequency and difficult to model and remove.

Spectral differences complicate matters by splitting the scene into a number of separate bands or images. The advantage of splitting an image up into spectral bands for spectral modeling often complicates ground control point selection. In an infrastructural setting, for example, some road systems, depending upon the reflectance characteristics of the surface material, appear well only on a particular band width. It is often necessary to look at all the bands in a data set to



build a complete picture of a road network. Obviously, in dealing with two sets of multiband data different in their band widths and number of bands, the matter of matching the road networks becomes even more complicated.

While spatial differences and platform differences between data inputs present more complex registration problems, the spectral problem of multiband image handling can be more easily solved.

## 2.0 PURPOSE

The objective of this study was to develop a simple technique designed to improve the accuracy of ground control point selection for geometric correction and registration of multiband and multisource imagery. The transformation used was selected such that it was currently available and a relatively standard accessory to most digital image processing packages. This was done in order that the technique presented here be immediately implementable by a majority of the user community.

Principal components analysis and its transformation was selected due to its relative simplicity and general availability. Jet Propulsion Laboratory's VICAR, ESL's IDIMS, and the Pennsylvania State University's ORSER system all have principal components options. Principal components is a technique whereby a new set of axes is defined for the data such that the first principal component or axis explains as much of the total variance as can be explained by any single variable or axis. The second principal component or axis explains as much of the remaining variance as can be explained by any axis orthogonal (uncorrelated) to the first. The third principal component continues this process and so on until the dimensionality of the data is exhausted (Merembeck and Borden, 1978). The effect is that most of the information inherent in the many spectral bands is combined or explained by one, two, or three of the principal components. The technique was used here to create improved single band imagery for geometric control point selection. In this example, the control points were used for image to image registration of two different types of remotely sensed data, airborne Thematic Mapper Simulator (TMS) and satellite Landsat Multispectral Scanner System (MSS) data. In this case a directed principal components analysis was used to focus the transformation around such infrastructural features as road networks, housing developments, and urban areas.

## 3.0 MATERIALS AND METHODS

The two images used in this study were a 7-channel TMS data set and a 4-channel Landsat MSS data set. Both sets of image data were collected over the city of Lancaster, Pennsylvania (Figure 1).

The TMS data is composed of 7 bands (0.45-0.52, 0.52-0.60, 0.63-0.69, 0.76-0.90, 1.55-1.75, 2.08-2.35, 10.4-12.5  $\mu\text{m}$ ), has a GIFOV of 30 x 30m and was collected from an airborne Learjet aircraft (ORI, 1982).

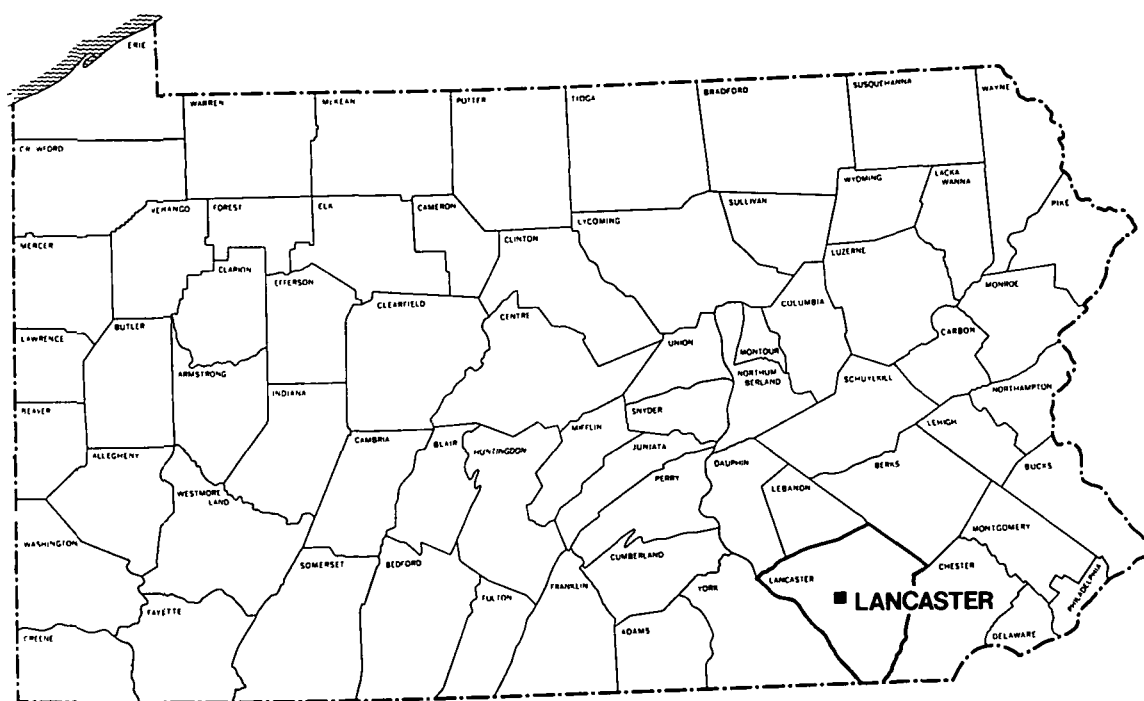


Figure 1: Study site, the city of Lancaster Pennsylvania

The TMS data suffered from several problems:

1. Considerable image distortions due to aircraft roll, pitch, and yaw were apparent in all bands.
2. A high frequency spatial distortion possibly due to aircraft and/or instrument jitter was also present in all bands.
3. Calibration problems and electronic noise also appeared in the imagery in the form of line striping and beat patterns.

The satellite data was Landsat-2 data tape ID# 21660-15011. It consisted of 4 spectral bands (0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-1.1  $\mu\text{m}$ ) and was resampled to create a 79 x 79 meter pixel. Some slight line striping was present in bands 1 and 2; however, for the most part the data quality was quite good.

The survey site consisted of a 236 km<sup>2</sup> area centered over the city of Lancaster, Pennsylvania. The equipment used to display, locate, transform, and analyze the image data consisted of a DeAnza image display system interfaced with ESL's IDIMS software implemented on an HP 3000 computer.

Ground control points were selected from USGS 7½ minute topographic quadrangles and color infrared photography. Nineteen widely scattered points were selected. The ground control points consisted of road intersections, street corners, and bridges. The ground control points selected were relatively stable over time and generally visible in all seasons of the year. Although the ground control points were visible on both data sets, it was necessary to view all of the MSS bands and four of the TMS bands (2, 4, 6, and 7) to accurately locate the exact position of the control points. This problem was found to be most prominent where highways composed of varying surface materials intersected or where a bridge passed over water.

Using the raw data sets the ground control points were located and their image coordinates fed into separate image mensuration files. Precise location of the control points using the raw data sets proved difficult. Many of the road intersections were visible but the actual road crossings could not be distinguished even using the three color (band) composite capability of the DeAnza image display.

When all 19 points were located as carefully as possible, the corresponding sets of mensuration data were input to an algorithm where a transformation matrix was generated using a least-squares fit method. The transformation was then applied to the TMS mensuration file and the resultant points were subtracted from the MSS mensuration file in order to determine residuals. All transformations for this study were carried out only to the first order for simplicity. The list of source (TMS) and destination (MSS) control points and their residuals for the raw data sets (Table 1) reveal errors in excess of 2.6 pixels in the x direction and 1.6 pixels in the y direction (residuals are measured in terms of Landsat 79 x 79m pixels).

In an effort to improve the accuracy of control point location, a directed principal component analysis was performed on the TMS and MSS imagery. A directed principal components transformation differs from general principal components in that instead of using statistics describing the general shape of the whole data swarm in four dimensional space, the statistics are derived from a subset of the data describing the target surface feature or features.

A simple polygon targeting training site selection and statistical calculation program was used to derive means and variance-covariance statistics for a specified region of the imagery. A polygon was defined that delineated only the more heavily urbanized areas supporting a large proportion of region's infrastructure. Only the urban areas, highway strip developments and high density residential areas were utilized for statistics definition. A set of correlation matrices were generated comparing the raw and transformed data sets for both the MSS and TMS data (figures 2 and 3).

MSS axes 1 and 2 had high correlations with MSS raw bands 1 and 2 (.5-.6  $\mu\text{m}$  and .6-.7  $\mu\text{m}$ ). This is not surprising since those bandwidths best describe rock, soil and manmade features. The relationship between the TMS transformed data (axes) and the TMS raw data is a little more complicated. With the exception of band 7 (thermal), no single raw band has appeared to contribute an inordinately large proportion of information to the principal components. Instead the principal components seem to be "extracting" nearly equal amounts of information from each of the raw bands.

The transformation based on the statistics for these areas reduced the data dimensionality and yielded imagery showing improved detail in the scene's urban and residential infrastructure (figures 4 and 5). TMS axis 1 portrayed roadways, intersections, and residential housing plans much more clearly than any other single band or 3-channel combination of the raw data.

The Landsat MSS data was also improved by the transformation. MSS axes 1 and 2 showed as much infrastructural detail as all four of the raw bands. The transformed data was used for the reselection of the previously defined ground control points.

The new control point coordinates for the TMS and MSS data were entered into mensuration files and the least-square fit method of determining the transformation matrix and residuals was repeated.

#### 4.0 RESULTS AND CONCLUSIONS

The geometric registration of remotely sensed data sets is an important factor in creating useful multisource composite imagery and input imagery for geographic information systems. Any improvement in geometric control point selection speed and accuracy is of great value to the process of GIS data entry.

FIGURE 2: CORRELATION MATRIX FOR MSS DATA  
PC AXES VS. UNALTERED BANDS

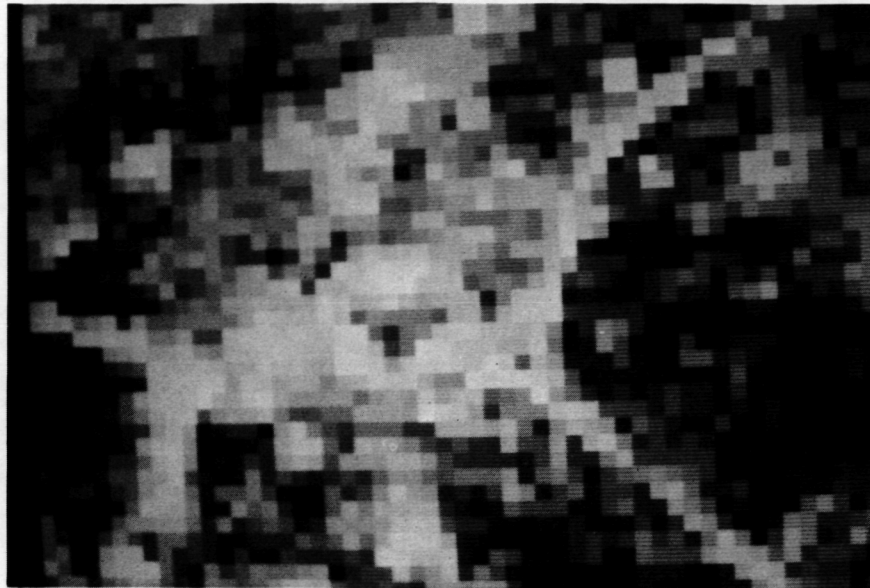
UNALTERED BANDS	PC Axes			
	1	2	3	4
1	.96	.31	-.01	.20
2	.98	.31	.01	-.16
3	.36	.96	.20	.01
4	-.17	.92	-.20	-.03

\* correlations are for urban targets only

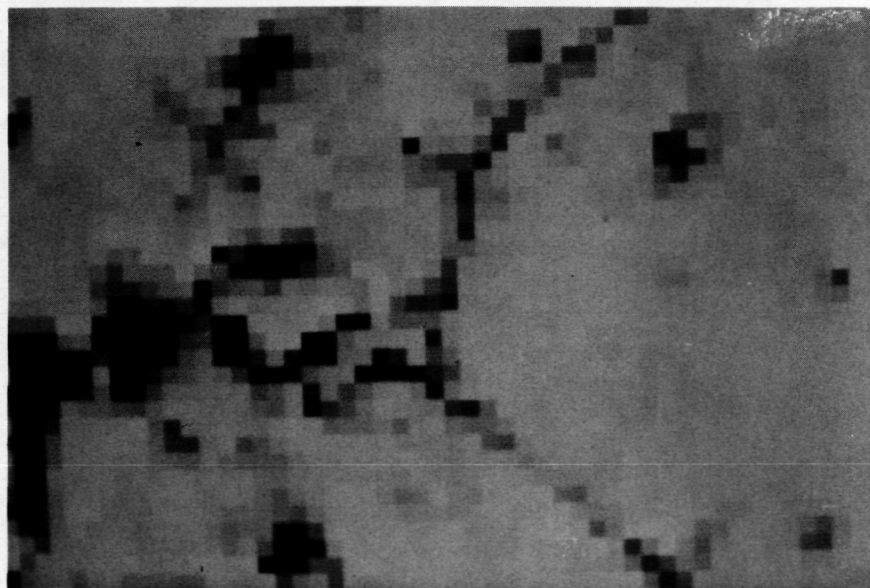
FIGURE 3: CORRELATION MATRIX FOR TMS DATA  
PC AXES VS. UNALTERED BANDS

UNALTERED BANDS	PC Axes						
	1	2	3	4	5	6	7
1	-.38	.60	-.63	.13	-.05	.34	-.21
2	-.50	.68	-.58	.21	.02	-.06	.07
3	-.47	.68	-.63	.27	.06	-.15	-.13
4	-.62	.24	.54	.03	.13	-.02	.03
5	-.62	.54	-.02	.61	-.11	-.04	-.01
6	-.42	.57	-.32	.51	.57	.01	-.05
7	.85	.64	-.07	-.02	.01	-.02	.01

\* correlations are for urban targets only

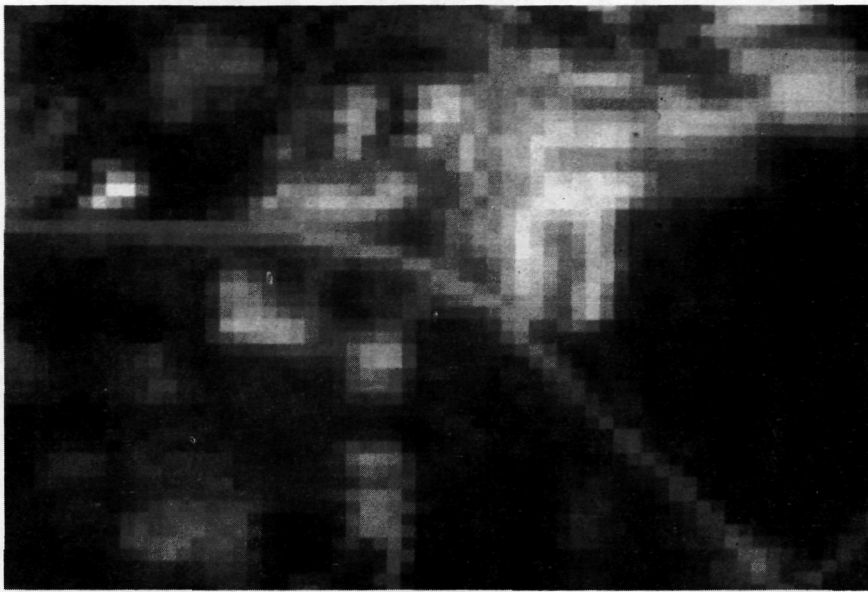


Raw MSS band 6

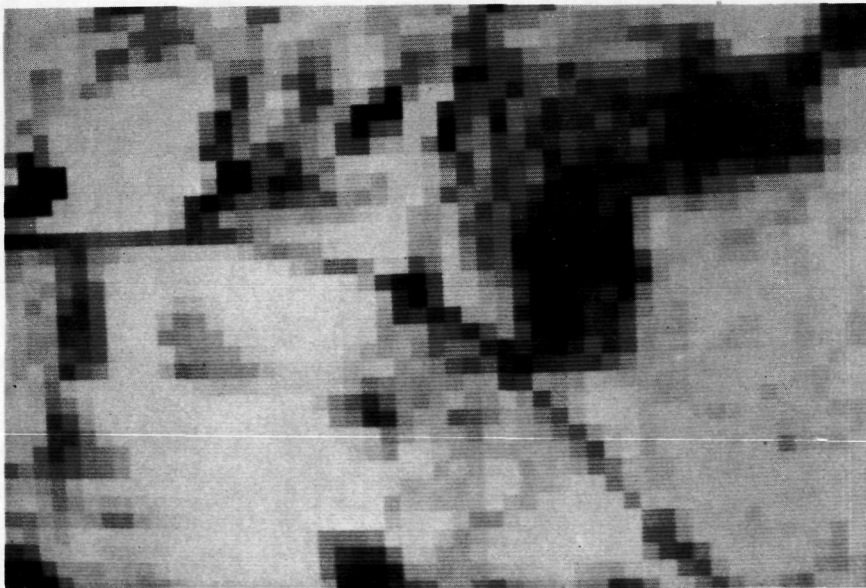


MSS principal component axis 1

FIGURE 4: Raw MSS imagery and MSS principal component axes 1 image depicting a road intersection



TM simulator band 4



TM simulator principal component axis 1

FIGURE 5: Raw TM simulator imagery and TM simulator principal component axis 1 depicting a road intersection

By showing greater detail of the infrastructural features, the principal components transformed image data allowed for a more accurate delineation of the ground control point locations. Road intersections could be more precisely located eliminating errors in their coordinate definition. The residuals from the least squares derived geometric transformation revealed that the control point reselection using the principal components enhanced data resulted in a combined 52 percent decrease in registration error. The x and y simple means of the residuals were reduced from 2.67 and 1.63 pixels for the raw data to 1.14 and 0.86 pixels for the principal components data (Tables 1 and 2). This represented a 57 percent reduction in error in x direction and a 47 percent reduction in locational error in the y direction.

In conclusion, a directed principal components analysis and its transformation can provide a valuable means for improving registration accuracy between remotely sensed imagery.

The technique provided single band imagery with improved target feature definition. The use of single band imagery expedited the ground control point selection process as it required less time to access, display and expand the single band imagery. The use of a single band display also eliminated potential resolution problems caused by color gun alignment within the display device. The technique is straightforward, consumes little time (30 minutes for complete analysis and transformation of a 1000 x 500 7-channel data set), and in this case provided a significant increase in accuracy.



TABLE 1: List of source, destination, and transformed coordinates and the residuals for control point matching using raw MSS and TMS data

GROUND CONTROL POINT	SOURCE (TMS)		DESTINATION (MSS)		TRANSFORMED		RESIDUALS	
	SX	SY	DX	DY	TX	TY	RX	RY
1 AX	194	474	217	213	216	212	1.4	.9
2 AR	77	670	176	288	165	296	10.7	-8.4
3 AM	224	706	230	314	235	313	-5.1	1.5
4 AN	139	550	191	245	192	245	-.6	.2
5 AD	94	459	171	208	168	205	2.6	2.7
6 AH	313	367	269	164	269	166	-.0	-2.3
7 AO	192	582	217	260	217	259	-.2	1.2
8 AI	255	520	246	233	245	232	.7	.7
9 AJ	249	613	242	275	245	272	-2.7	2.6
10 AK	263	586	249	262	251	261	-1.6	1.2
11 AL	324	550	280	246	278	245	1.6	.5
12 AG	328	478	281	214	279	214	2.4	-.4
13 AC	151	449	195	203	195	201	.1	1.8
14 AE	191	420	209	188	213	189	-4.0	-.8
15 AB	240	432	238	195	236	194	1.7	.8
16 AI	292	441	262	198	261	198	1.1	-.3
17 AQ	177	691	209	306	213	306	-3.7	.1
18 AF	377	320	301	143	298	146	3.0	-3.3
19 AS	126	379	174	172	182	171	-7.6	1.2
RESIDUAL MEANS	=	2.67	1.63					

TABLE 2: List of source, destination, and transformed coordinates and the residuals for control point matching using single band PC transformed MSS and TMS data

GROUND CONTROL POINT	SOURCE (TMS)		DESTINATION (MSS)		TRANSFORMED		RESIDUALS	
	SX	SY	DX	DY	TX	TY	RX	RY
1 AR	99	665	171	295	172	297	-1.1	-1.7
2 AM	226	706	230	314	232	314	-2.4	.3
3 AS	126	379	181	173	183	172	-2.2	1.0
4 AF	377	320	301	143	302	145	-.6	-1.6
5 AQ	177	691	209	306	209	308	-.2	-1.5
6 AI	292	441	262	198	262	198	-.1	.1
7 AB	240	432	238	195	237	194	.6	.7
8 AE	191	420	214	189	214	189	-.2	-.4
9 AC	151	449	195	203	195	202	-.4	.7
10 AG	328	478	281	214	279	214	1.6	.3
11 AL	324	550	280	246	278	245	2.1	.9
12 AK	263	586	249	262	249	261	-.2	.8
13 AJ	249	613	242	275	243	273	-.8	1.9
14 AI	255	520	246	233	245	233	.9	.5
15 AO	192	582	217	260	216	260	1.4	.0
16 AH	313	367	269	164	272	165	-2.6	-1.5
17 AD	94	459	171	208	169	207	2.5	.9
18 AN	139	550	191	245	190	246	.6	-1.4
19 AX	194	474	217	213	216	213	1.1	.1
RESIDUAL MEANS	=	1.14	.856					

## BIBLIOGRAPHY

- Billingsley, F. C., 1982. Revised Edition of The Manual of Remote Sensing, Chapter 17, American Society of Photogrammetry.
- Jayroe, R. R., Jr., 1976. Nearest Neighbor, Bilinear Interpolation and Bicubic Interpolation-Geographic Correction Effects on Landsat Imagery, NASA Technical Memorandum TMX-73345, Marshall Space Flight Center.
- Logan, T. L., and A. H. Strahler, 1979. The Errors Associated with Digital Resampling of Landsat Forest Imagery for Multidate Registration, Eighth Annual Remote Sensing of Earth Resources Conference, The University of Tennessee Space Institute, Tullahoma, Tennessee.
- Merembeck, B. F., and F. Y. Borden, 1978. Principal Components and Canonical Analysis for Reduction of the Dimensionality of Large Data Sets. Technical Report 5-78, Office for Remote Sensing of Earth Resources, The Pennsylvania State University, University Park, Pennsylvania.
- ORI. Renewable Resources Thematic Mapper Simulator Workshop, Technical Report 1907k ORI, Silver Spring, Maryland.

# GROUND TRUTH SAMPLING AND LANDSAT ACCURACY ASSESSMENT

by

Jon W. Robinson

Computer Sciences Corporation

Fred J. Gunther

Computer Sciences Corporation

William J. Campbell

Goddard Space Flight Center

## INTRODUCTION

The work reported in this paper was supported by a contract (The Power Plant Siting Study) from the Nuclear Regulatory Commission to the National Aeronautics and Space Administration, Goddard Space Flight Center. The work was carried out by government and contractor personnel at Goddard Space Flight Center in cooperation with the Nuclear Regulatory Commission and Pennsylvania Power and Light Company.

The purpose of the study was to compare the cost and accuracy of various remote sensing data types and processing procedures for updating Geographic Information Systems (GIS). This paper reports a portion of the work carried out under that contract. A complete report of the work carried out under the contract will be submitted to the Nuclear Regulatory Commission at the end of the contract period and will be available to the public from the Nuclear Regulatory Commission.

The key factor in any accuracy assessment of remote sensing data is the method used for determining the ground truth, independent of the remote sensing data itself. This paper will describe the sampling and accuracy procedures developed for the Power Plant Siting Study.

The purpose of the sampling procedure was to provide data for developing supervised classifications for the two study sites and for assessing the accuracy of that and the other procedures used. The purpose of the accuracy assessment was to allow the comparison of the cost and accuracy of various classification procedures as applied to various data types.

There were two study sites, one centered on the city of Lancaster, Pennsylvania and the other centered on the Susquehanna Steam (nuclear) Generating Plant near Berwick, Pennsylvania. The methods described here were used at both sites, but only the results from the Berwick site will be presented here. The final report to the Nuclear Regulatory Commission will contain the results from both sites.

Each site contained 400 square miles, 20 miles on a side. Both sites were within the Pennsylvania Power and Light Company's service area and were covered by that company's Environmental Land Use Data System (ELUDS) data base (a geographic information system). The data base includes a variety of data types, including land cover, geology, slope, infrastructure, and historic sites.

## METHODS

In this section, the materials used and the methods employed for both the sampling procedure and the accuracy assessment procedure will be presented. The sampling and accuracy procedures involved the use and merging of several data types. These included Landsat Multispectral Scanner (MSS) data, Thematic Mapper Simulator (TMS) and low altitude aerial photography which was digitized for further manipulation by computer. All of these data were registered to United States Geological Survey 7.5-minute maps so they would be congruent with each other. The results of a ground survey were then combined with the previous data to provide estimates of the accuracy of the two types of classifiers used on the MSS and TMS data. Since the study area was too large to be completely surveyed, a sampling procedure was developed.

### Sampling Methods

The goal of the sampling procedure was to generate as many ground truth pixels per given amount of effort as possible, yet maintain a statistically valid procedure. The sampling procedure chosen was cluster sampling (Cochran, 1977). This allowed areas to be chosen at random and a large number of pixels to be identified in each chosen area.

The areas were chosen by taking United States Geological Survey (USGS) 7.5-minute quadrangle maps of the study site and picking points at random from selected quadrangles. Because of time constraints, a contiguous group of maps within the study area was selected. That group of maps included the Susquehanna Steam Generating Plant.

The vertical and horizontal borders of each map were marked at one inch intervals. Pairs of two-digit random numbers were then taken from a random number table (Rohlf and Sokal, 1969) to select pairs of horizontal and vertical tick marks from the edges of the maps. If a two-digit number was beyond the range of the tick marks, another two-digit number would be chosen until one within the range was selected. Each pair of tick marks identified a centroid of a one-inch-by-one inch square on the map. Due to the dense road network, each square selected on the map was crossed by or closely approached by at least one road. Each site so selected was then visited with a survey crew provided by Pennsylvania Power and Light Company. Table 1 lists the name of each quadrangle selected and the approximate latitude and longitude of each site visited within that quadrangle.

TABLE 1

Latitude and Longitude of Ground Truth Sample Areas

Quadrangle	#	Latitude	Longitude
Shickshinny	6	41 9.3 N	76 9.0 W
"	4	41 10.0 N	76 10.9 W
"	1	41 10.7 N	76 12.2 W
"	2	41 13.3 N	76 10.0 W
"	3	41 13.3 N	76 11.0 W
"	5	41 13.0 N	76 14.1 W
"	7	41 12.6 N	76 14.8 W
Stillwater	14	41 9.4 N	76 18.2 W
"	16	41 2.9 N	76 19.4 W
"	15	41 8.2 N	76 18.0 W

On arriving at a site, landmarks that would show up on low altitude aerial photography were identified. Then the location of field boundaries and the boundaries between landcover types were measured relative to the landmarks. Detailed notes on the crop types and landcover types surveyed were taken along with 35mm. photographs on Kodachrome and Infrared Aero Ektachrome. The Infrared Ektachrome pictures were taken so that the observations obtained on the ground could be compared with low altitude color Infrared photography and Infrared photography taken by the Thematic Mapper Simulator flight.

The original plan was to have the low altitude aerial photography performed on or close to the date of the field work which was during the last week of August 1981 and to have this coincide with the flight of the Thematic Mapper Simulator (TMS). The low altitude photography was being provided by a subcontractor for Edgerton Gearson & Greer Corporation (EG&G) for the Nuclear Regulatory Commission on a separate contract. Because of contracting delays, the flight was not made until

the 25th of September 1981. The Thematic Mapper Simulator (TMS) flight was being flown by National Aeronautics & Space Administration National Space Technology Laboratories (NSTL) in Mississippi. Although the field work was undertaken with the understanding that NSTL would make the TMS flight during the ground-truth field work, it was in actuality not flown until the 12th of October.

The low altitude aerial photography was digitized by the University of California Santa Barbara on a (subcontract from EG&G) into three digital images for each frame. Each digital image was filtered by the appropriate red, green or blue filter so that the color information content of the original color infrared photograph would be retained. Each frame of digitized photography was entered into the Interactive Digital Image Manipulation System (IDIMS) on a HP3000 computer. Each frame that covered one of the ground-truth study sites was then registered to the 7.5-minute quadrangle map in which it occurred. The registration was to within 15 meters, which is the accuracy limit of the 7.5-minute quadrangle maps.

The registered images were then displayed on a color raster display using the IDIMS programs; and the boundaries of the landcover types were drawn in and the polygons thus generated labeled using the data collected during the ground-truth collection field trip. Because all of the remote-sensing images were registered to the same 7.5-minute maps, the identity of any pixel falling within one of the ground-truth polygons could be determined. Thus, the accuracy of the classifications generated by the various processing methods could be determined for each type of data used by counting the number of pixels of known ground cover that were correctly labeled by a classification.

#### Accuracy Methods

For the accuracy assessment, the identity of pixels falling within the ground truth polygons and urban-area polygons (which were photointerpreted) were compared with the classification labels produced by a particular classification method. The two primary methods of classification used were maximum likelihood and cluster analysis with the ISOCLS routine in the IDIMS system.

The maximum likelihood classifier required that statistics, sample mean vectors and sample variance-covariance matrices be generated for each landcover type. Half of the ground truth sites were used to generate these statistics and the other half were used to estimate the accuracy of the method.

Theoretically, one could use the pixels used to generate the maximum likelihood decision rule to estimate its accuracy. This estimate of the accuracy would only be unbiased if the sample used to generate the classification was unbiased. Therefore it is best to use an independent sample of pixels, if that is possible, to test the accuracy of a maximum likelihood classifier. The practice of using the classifier to classify the pixels which generated it and then using the accuracy of that classification to estimate accuracy of the classifier is called back classification. A close agreement between accuracy estimates from back classification and from a classification of an independent sample of pixels of known identity indicates that the two samples are less likely to have been drawn in a biased manner from the population of pixels and that more faith can be placed in the estimates so derived.

Thus to check for bias in selecting which sites would be used for generating the classification and which sites would be used for accuracy determination, the back classification accuracy was determined for the training site pixels as well as for an independent sample of pixels.

Because the ground-truth sites had been broken into two groups for testing the accuracy of the maximum likelihood classification, the accuracy of the ISOCLS classifications were estimated by comparing the accuracy for each group of ground-truth sites separately. This provided two independent estimates of the accuracy for each ISOCLS classification.

A table like table 2, was generated from a CONTABLE (an IDIMS program) run on each classification. The values in these tables were then used to calculate the following estimates: the probability that a pixel is correctly classified; the probability that a pixel belonging to class 1 is classified into class 1, and the probability that a pixel classified as class 1 is in fact a member of class 1.

Table 2 shows the unweighted procedure for calculating accuracy figures. This means that the number of pixels in each category are in proportion to their frequency in the ground truth polygons. Because urban areas were photointerpreted, the relative frequency of those pixels in the accuracy assessment procedure were greater than their relative frequency in the image being classified. If the accuracy figures were adjusted to the relative frequency of each category of pixel in the image being classified, then they would be weighted (or a weighted accuracy assessment).

It has been pointed out (Chrisman, 1980) that simple accuracy figures, by themselves, may be misleading. A better measure of how well a classifier is performing would be the percentage improvement over a random classifier based on the



relative frequencies of the classes. The kappa statistic (Everitt, 1968) provides such a measure. Using the frequency of pixels in each class in the ground-truth polygons to calculate the expected frequencies for a random classifier, the kappa statistic was calculated for each data type and classification procedure.

TABLE 2

## Accuracy Calculations

	Classifier Label					Number Belonging To Each Class
	1	2	3	...	M	
True Label						
1	$m_{11}$	$m_{12}$	$m_{13}$	...	$m_{1M}$	$m_{1\cdot}$
2	$m_{21}$	$m_{22}$	$m_{23}$	...	$m_{2M}$	$m_{2\cdot}$
3	$m_{31}$	$m_{32}$	$m_{33}$	...	$m_{3M}$	$m_{3\cdot}$
.	.	.	.	...	.	.
.	.	.	.	...	.	.
.	.	.	.	...	.	.
M	$m_{M1}$	$m_{M2}$	$m_{M3}$	...	$m_{MM}$	$m_{M\cdot}$
Number Classified As	$m_{\cdot 1}$	$m_{\cdot 2}$	$m_{\cdot 3}$	...	$m_{\cdot M}$	$m_{\cdot \cdot}$

$$\text{Total Pixels Checked} = TP = \sum_{i=1}^M \sum_{j=1}^M m_{ij} = m_{\cdot \cdot}$$

$$\text{Total Pixels Correct} = TC = \sum_{i=1}^M m_{ii}$$

Probability that a pixel  
in the sample is correctly  
classified.  $P_{cc} = TC/TP$

Probability that a pixel  
classified as class  $i$  is  
a member of class  $i$ .  $P_{ci} = m_{ii} / \sum_{j=1}^M m_{ji} = m_{ii} / m_{\cdot i}$

Probability that a pixel  
that is a member of class  $i$   
is classified as class  $i$ .  $P_{ic} = m_{ii} / \sum_{j=1}^M m_{ij} = m_{ii} / m_{i\cdot}$

## RESULTS

The results of the sampling can only be presented in terms of an analysis of the accuracy figures. Table 3a gives the results of the unweighted accuracy calculations based on the maximum likelihood classification of the independent sample of pixels of known identity for both the MSS and TMS images. Table 3b gives the results of the unweighted accuracy calculations based on the maximum likelihood classification of the pixels used to generate the classification functions (back classification).

Tables 4a and 4b present similar results for the unsupervised method (cluster analysis) of classification. Because the classes are not predefined as in the supervised method (maximum likelihood) the analyst must assign names to the classes generated by the clustering algorithm. This led to the merging of several ELUDS landcover classes into more general categories. The merged ELUDS classes are identified by the numbers associated with each landcover name in tables 4a and 4b.

It should be noted that those categories that have small samples for the training sets, i.e. the N columns in table 3b, have low accuracies. Beyond this, the results for the accuracy assessment based on the back classification are not very different from those based on the independent sample. The small pixel counts for the landcover class "barren land" in the unsupervised classification do not provide an accurate estimate of the probabilities for that class.

There is little difference between the probabilities of correct classification for the different classification methods. The primary difference is in the number of classes that can be differentiated. The kappa statistic also reflects this situation.

The overall quality of the classifications based on the TMS data are better for all of the classification procedures and assessment data sets. Since the quality of the TMS data was very bad it contained a large amount of noise, the quality of classifications based on real Thematic Mapper (TM) data should be better.

TABLE 3A  
BERWICK  
ACCURACY ASSESSMENT MAXIMUM LIKELIHOOD CLASSIFIER  
(INDEPENDENT SAMPLE)

ELUDS CODE	FREQ.*	MSS			TMS		
		N	P <sub>ci</sub>	P <sub>ic</sub>	N	P <sub>ci</sub>	P <sub>ic</sub>
1 URBAN	.0537	1572	.882	.803	6091	.930	.953
2 BARREN LAND	.0375	68	.186	.353	116	.019	.017
3 AGRICULTURAL	.3225	568	.418	.563	2110	.696	.667
5 TREE PLANTAT.	.0018	7	.0	.0	27	.039	.111
7 CONIF. FOREST	.0084	18	.0	.0	77	.055	.182
9 DECID. FOREST	.3852	1490	.780	.590	2694	.675	.591
11 MIXED FOREST	.1603	138	.071	.094	512	.073	.060
13 SCRUB LAND	.0048		NONE			NONE	
14 MEADOW	.0009	15	.0	.0	50	.0	.0
15 FORESTED WETL	.0099	26	.023	.115		NONE	
16 UNFOREST WETL	.0000		NONE			NONE	
99 WATER	.0148	212	.864	.962	843	.840	.890

$$P_{cc} = .6578$$

$$P_{cc} = .7669$$

$$KAPPA = .5108$$

$$KAPPA = .6590$$

\*FREQ. - The frequency of each ELUDS data type in the entire 400 square mile Berwick study site.

N - The counts of pixels of each ELUDS landcover type in the ground truth polygons used for the independent accuracy assessment.

TABLE 3B  
BERWICK  
ACCURACY ASSESSMENT MAXIMUM LIKELIHOOD CLASSIFIER  
(BACK-CLASSIFICATION)

ELUDS CODE	FREQ.*	MSS			TMS		
		N	P <sub>ci</sub>	P <sub>ic</sub>	N	P <sub>ci</sub>	P <sub>ic</sub>
1 URBAN	.0537	1567	.951	.678	6305	.977	.908
2 BARREN LAND	.0375	104	.121	.106	61	.750	.443
3 AGRICULTURAL	.3225	285	.281	.537	1070	.688	.827
5 TREE PLANTAT.	.0018	6	.231	1.000	23	.188	.826
7 CONIF. FOREST	.0084	83	.619	.157	347	.441	.478
9 DECID. FOREST	.3852	1125	.732	.762	2439	.832	.801
11 MIXED FOREST	.1603	24	.066	.333	97	.179	.433
13 SCRUB LAND	.0048		NONE			NONE	
14 MEADOW	.0009	12	.240	.500	40	.440	.825
15 FORESTED WETL	.0099	10	.063	.600		NONE	
16 UNFOREST WETL	.0000		NONE			NONE	
99 WATER	.0148	96	.929	.958	349	.795	.943

$P_{cc} = .6685$

$P_{cc} = .8553$

KAPPA = .4906

KAPPA = .7726

\*FREQ. - The frequency of each ELUDS data type in the entire 400 square mile Berwick study site.

N - The counts of pixels of each ELUDS landcover type in the ground truth polygons. This is also the sample size for each class's training set.

TABLE 4A

BERWICK  
ACCURACY ASSESSMENT UNSUPERVISED CLASSIFICATION

(INDEPENDENT SAMPLE)<sup>1</sup>

LAND COVER	FREQ. *	N	MSS		N	TMS	
			P <sub>ci</sub>	P <sub>ic</sub>		P <sub>ci</sub>	P <sub>ic</sub>
1 URBAN	.0537	1572	.953	.502	6076	.988	.706
2 BARREN LAND	.0375	68	.289	.191	114	.073	.491
3 + 14 AGRICUL.	.3234	583	.369	.877	2160	.583	.787
5 + 7 + 9 + 11 + 15 FOREST	.5656	1679	.855	.846	3313	.836	.912
99 WATER	.0148	212	.964	.892	838	.925	.952

$$P_{cc} = .7103$$

$$P_{cc} = .7892$$

$$KAPPA = .5639$$

$$KAPPA = .6802$$

<sup>1</sup>The subtitle Independent sample is used for identification purposes only. The unsupervised classification procedure does not use training sites.

\*FREQ. - The frequency of each ELUDS data type in the entire 400 square mile Berwick study site.

N - The counts of pixels of each landcover type in the ground truth polygons used for the independent accuracy assessment.

TABLE 4B  
BERWICK  
ACCURACY ASSESSMENT UNSUPERVISED CLASSIFICATION  
(BACK-CLASSIFICATION)<sup>1</sup>

LAND COVER	FREQ.*	MSS			TMS		
		N	P <sub>ci</sub>	P <sub>ic</sub>	N	P <sub>ci</sub>	P <sub>ic</sub>
1 URBAN	.0537	1567	.967	.318	6292	.998	.512
2 BARREN LAND	.0375	104	.018	.019	58	.028	.448
3 + 14 AGRICUL.	.3234	297	.190	.761	1107	.281	.703
5 + 7 + 9 + 11 + 15 FOREST	.5656	1248	.750	.851	2873	.738	.869
99 WATER	.0148	96	1.000	.865	336	.898	.917

$$P_{cc} = .5652$$

$$P_{cc} = .6407$$

$$KAPPA = .3036$$

$$KAPPA = .3671$$

<sup>1</sup>The subtitle back-classification is used for identification purposes only. The unsupervised classification procedure does not use training sites.

\*FREQ. - The frequency of each ELUDS data type in the entire 400 square mile Berwick study site.

N - The counts of pixels of each landcover type in the ground truth polygons used for the training of the maximum likelihood.

## DISCUSSION

The results of the accuracy assessment of the supervised classification indicate that there was no strong bias in the sampling procedure. The low accuracies for certain categories may be due to either the similarities in their spectral reflectivities or the small samples used to characterize their spectral reflectivities. At an intuitive level, it is easy to understand how the various forest landcover types would be spectrally confusing. The causes of confusion between the other classes are not so obvious.

One remedy for the small sample sizes of certain categories would be to use a stratified sampling procedure (Cochran, 1977), where the strata would be the landcover categories. This would allow for adequate sample sizes for all but the rarest categories. There is one requirement for this procedure that makes it more difficult to carry out. That is, a landcover map of the area must already be available. It does not have to be perfect, but it must be sufficiently accurate so that the majority of the field checks are made in the correct categories.

A further problem with cluster sampling is that neighboring pixels are used for the training set pixels and for the accuracy assessment pixels. Studies by a variety of authors have shown that the spectral characters of the pixels are spatially autocorrelated. It is also clear that other characteristics may be spatially autocorrelated. Since one of the basic assumptions behind the estimation procedures used is that the observations are statistically independent, the confidence bounds of the quantities presented here can not be reliably determined. Further, because of theoretical considerations it may be that the classifications themselves would be quite different if the autocorrelation in the spectral values of neighboring pixels were removed.

The overall accuracies of the two classification procedures do not differ much between themselves when compared with the variation within a procedure. The prime differences are that in the supervised classification, the classes are defined in advance and that in the unsupervised classification, the classes are assigned names on an adhoc basis. The success of the adhoc assignment of class identities by the skilled analyst are vindicated by the small differences between the supervised classification and the unsupervised classification accuracies.

A major consideration in choosing which classification procedure will be used in a study will be cost. The cost to properly execute a supervised classification is considerably greater than the cost to properly execute an unsupervised classification. In many situations, where the classes of



landcover that are to be distinguished are coarse, the unsupervised methods are the most efficient. In those situations where a statistically rigorous procedure is required and where many categories must be distinguished, the extra cost of the supervised procedure may be justified.

The accuracies achieved by both classification methods were consistently better with the TMS data than with the MSS data. This was inspite of the fact that the TMS data was very noisy and required both geometric and spectral correction for the bow tie effect. This indicates that the increased spectral and spatial resolution provide for a consistently more accurate classification. The results with real Thematic Mapper data should be much better than the results presented here.

A more detailed analysis of the data developed in this study should provide a better understanding of the results presented here. Such analysis could look at the trade off between noise in individual sensor channels and greater spectral and spatial resolution. Such analysis could also examine the effects of autocorrelation on all aspects of a classification procedure: the classification, and the accuracy estimates.

#### CONCLUSION

The sampling design and the associated accuracy assessment presented above indicate that Thematic Mapper data should provide consistently better classification results than the old Multispectral Scanner data of Landsat 1, 2 and 3. In addition it appears that the choice of a classification procedure will depend on the purposes to which the classification will be put and the resources available to execute it. In a supervised classification the sampling procedure by which ground truth is obtained will be dictated by the requirements of the particular study. If the accurate classification of rare classes is not of great importance, than cluster sampling may prove quite efficient. However, other sampling procedures should be considered when rare classes are important and the necessary ancillary information is available.

#### LITERATURE CITED

Anonymous. 1981. IDIMS Functional Guide Volume 1. Technical Manual ESL-TM705. ESL Incorporated. Sunnyvale, California. viii + 716p

Chrisman, Nicholas R. 1980. Assessing Landsat accuracy and correcting for missclassification errors. Unpublished Manuscript. 18p.

Cochran, William G. 1977. Sampling Techniques, third edition. John Wiley & Sons, xvi + 428.

Everitt, B. S. 1968. Moments of the statistics kappa and weighted kappa. The British Journal of Mathematical and Statistical Psychology. 21:97-103.

Rohlf, F. James and Robert R. Sokal. 1969. Statistical Tables. W. H. Freeman and Company. xi + 253.

LANDSAT DIGITAL DATA BASE PREPARATION FOR THE  
PENNSYLVANIA DEFOLIATION APPLICATION PILOT TEST

R. G. McLeod and A. Zobrist

California Institute of Technology  
Jet Propulsion Laboratory  
4800 Oak Grove Drive  
Pasadena, California 91109 U.S.A.

ABSTRACT

Landsat digital data are convenient and adaptable sources of data to incorporate as a base in a geographic information system. These data are readily convertible to various map projections and scales and provide the user/analyst with a format similar to that of an aerial photograph. Certain properties associated with the data, however, inhibit widespread use. The framing convention of the Landsat sensor does not lend itself well to imaging entire states or provinces at the required resolution cells. For large areas, digital Landsat data must be geometrically corrected to a standard map projection and then mosaicked.

A Landsat digital mosaic data base for the State of Pennsylvania was prepared for use in the development of an automated system to annually estimate the extent and severity of Gypsy Moth defoliation of hardwood forests. The techniques for detecting the defoliation and development of a Geographic Information System (GIS) to assess damage is being developed jointly by NASA/Goddard Space Flight Center and Pennsylvania State University using the JPL prepared mosaic base. JPL processing involved the use of ground control points from the Master Data Processor (MDP) for planimetric control, resampling of the Landsat data to 57 x 57 meter pixels, realignment to north, and reprojection to the Universal Transverse Mercator (UTM) projection in UTM zones 17 and 18. The completed mosaic for each UTM zone was subdivided into 1 degree of latitude by 2 degrees of longitude quadrangles for easy data handling.

Consideration is given to the issues of mapping standards, sensor and spacecraft platform characteristics, and their implication to geographic information systems operation. Methods for obtaining measures of accuracy for Landsat mosaics are reviewed.

1. INTRODUCTION

Since its introduction from Europe into Massachusetts in the late 1860's, the Gypsy Moth Lymantria dispar (L.), has repeatedly defoliated hundreds of thousands of acres of forest. The mature Gypsy Moth caterpillar is about 2 to 3 inches in length, and as many as 30,000 of these caterpillars can infest a single tree. Each caterpillar can consume up to ten small leaves a day.[1] Over the past ten years, the State of Pennsylvania has attributed the loss of

\$32 million dollars worth of timber resources to this pest. The insect does not kill the tree immediately, but after prolonged infestations over several years the tree is destroyed. While the natural spread of the Gypsy Moth is slow, it can move rapidly because of its ability to hitchhike with people traveling through infested areas.

In order to plan appropriate pest management activities, resource managers must continually monitor the movements and damage caused by this insect. Over large geographic areas, conventional methods of surveillance such as field site visits and large-scale aerial photography are prohibitive to use because of cost and time. Alternative methods of assessment must be developed that are inexpensive, timely, and mesh well with current practices.

Developing new assessment methods for Gypsy Moth infestations is the goal of the Pennsylvania Defoliation Applications Pilot Test (APT), a joint study by Goddard Space Flight Center/NASA and Pennsylvania State University. These new methods being developed are to be transferred to the Pennsylvania Division of Forest Pest Management, Bureau of Forestry, for implementation to operational use.

The basic procedure is to utilize multi-date Landsat imagery to monitor the infestations.[2] An image is acquired for an area prior to infestation, and it is classified, using computer aided analysis techniques, to identify the extent of forest cover versus non-forest cover. After insect damage, a second image of the same area is obtained and it is digitally overlaid onto the forest cover map derived from the initial image. Forested areas exhibiting defoliation can then be identified and tabulated. Acreage counts and estimates can be generated and abatement procedures or strategies developed.

While Landsat is a convenient and relatively inexpensive source of data, certain properties associated with the data present problems. The framing convention of the Landsat sensor does not lend itself well to imaging entire states in a single scene. To increase the utility of the data, the Landsat frames must be geometrically corrected to a standard map projection and then mosaicked.

Goddard Space Flight Center, the lead center in this project, initiated a contract with Jet Propulsion Laboratory to prepare a Landsat digital mosaic of the State of Pennsylvania that will be used to address this problem. Three separate mosaics were prepared for the task: (1) an early date mosaic prior to defoliation; (2) the derived forest /non-forest cover map mosaic, and (3) a late date mosaic after defoliation.[3]

## 2. EARLY DATE MOSAIC

The Landsat data tapes used for the mosaic prior to defoliation were delivered to JPL by Goddard Space Flight Center. Goddard had originally ordered the scenes from EROS Data Center in order to proceed in a parallel effort with other aspects of the project. Table I depicts the Landsat frames

used in this mosaic, and Figure 1 shows the individual footprint of each scene for the state.

All processing performed at JPL utilized the Image Processing Laboratory (IPL). IPL hardware resources include an IBM 370/158 with 8 megabytes of memory, eight tape drives, and 3700 megabytes of on-line disk storage. The disk storage consists of CDC 3350 high speed disk drives. Image displays include a Ramtek display system that accomodates 6 bit black and white imagery up to 640 x 512 elements. The other system used consists of two COMTAL display units. A COMTAL 8003 system provides 512 x 512 element resolution for 8 bit color images and includes graphics planes and trackball cursors. A COMTAL 1024 system provides capability to display black and white images at a 1024 x 1024 element resolution.

The IPL also maintains a complete library of over 300 special purpose image processing applications programs. The system in use is the Video Image Communication and Retrieval (VICAR) and the Image Based Information System (IBIS) developed at JPL.[4,5]. This system is available from COSMIC for a nominal charge.[6]

Table I.

PATH	ROW	SCENE IDENTIFICATION	LOCATION NAME	DATE
19	31	21267-15031	Titusville	July 12, 1978
19	32	21267-15034	Steubenville	July 12, 1978
18	31	2600-15094	Warren	September 13, 1976
18	32	2600-15100	Pittsburgh	September 13, 1976
17	31	30478-15123	Williamsport	June 26, 1979
17	32	30208-15141	Harrisburg	September 29, 1978
16	31	21660-15005	Scranton	August 9, 1979
16	32	2544-15001	Lebanon	July 19, 1976
15	31	30170-15020	Poughkeepsie	August 22, 1978
15	32	30098-15013	Trenton	June 11, 1978

## 2.1 Logging the Initial Scenes

The Landsat data were initially logged to be compatible with the VICAR format and system requirements. The logging consists of a series of separate steps depending upon the type of data ordered. Since February 1979, imagery

processed by EROS is in band sequential format with major geometric corrections. If the data are processed prior to that date, the data are in band interleaved by pixel pairs with no geometric corrections performed. Typical of almost all applications involving this type of imagery, it was necessary to select acquisition dates spanning over a long period of time to obtain the most cloud free coverage possible. Hence, it was necessary to use both band sequential and band interleaved formats as basic data input for the task.

Imagery processed since February 1979 is fairly easy and inexpensive to log because no geometry changes are necessary, at least in the first phases of the mosaicking process. Extraneous engineering files are stripped off and a VICAR label attached to the image files to be used by subsequent VICAR modules. The uncorrected data, band interleaved by pixel pairs, require extended effort and expense to produce a data format suitable for the VICAR mosaicking process. Nominal geometric and radiometric corrections are made, in addition to dis-interleaving the image strips. Nominal corrections include removal of earth rotation induced skew, panorama effect, and mirror scan velocity profile (MSVP) compensation. The pixel size at this stage of the processing is the IFOV of 57 by 79 meters.

Every effort was made to obtain the clearest possible imagery during the growing season. There were a few problems with some individual scenes with respect to haze and overcast. The net effect of the haze is to reduce the variance in the scene while increasing the brightness. This poses particularly difficult problems when trying to match scenes radiometrically, and also when trying to extend multispectral signatures from one part of a scene to another part of the same scene.

## 2.2 Map Base

The Universal Transverse Mercator UTM Projection was chosen as the mapping base for the mosaic. It was decided to maintain a pixel size of 57 meters by 57 meters because of the IFOV sampling interval along the Landsat scan line. Selection of a 50 meter pixel size would have allowed the data to be selected from the UTM grid more conveniently, but would also have increased the amount of data to be processed while not increasing the information content.

The State of Pennsylvania covers about 6 degrees of longitude, large enough to encompass one UTM zone. Unfortunately, the state straddles a UTM zone boundary which bisects the state into a western and eastern zone, Zone 17, and Zone 18, respectively. To preserve map projection properties and to provide consistency with subsequent data sets to be registered to the Landsat mosaic data base, two separate mosaics were constructed, one for each zone. Coverage of the entire state with Landsat data can be met with ten scenes, but because of the two projection zones, six scenes were mosaicked for each zone, with the two central scenes contributing data to each zone. In effect, two six-frame mosaics were constructed for the task.

The mapping grid was configured so that the imagery would be resampled to the selected scale of 57 meters and rotated north assuring the data scan lines would be aligned east-west relative to the mapping grid. The advantages of

this technique are fairly straightforward. First, the data are displayed in a familiar fashion with north at the top, and second, map quadrangles can be extracted from the data base with a minimum of wasted storage space that results from rotation.

### 2.3 Planimetric Control

Planimetric control for remotely sensed imagery in a mosaicking context can be obtained in several ways. If the exposure or acquisition time of the scene is short enough, such as in a framing type sensor, calibration and control of the data using spacecraft ephemeris information is often sufficient. Since the scene acquisition time for the Landsat image is on the order of 27 seconds, and because it is a scanner type sensor, it is necessary to incorporate known geodetic points on the surface of the earth. Information obtained from the Control Point Library Building System (CPLBS) was used to provide planimetric control to each Landsat scene as each fits into the mosaic.[7]

The information from the CPLBS consists of a 32 pixel by 32 pixel image chip containing a geographic feature, e.g., a road intersection or river bend, as well as the latitude and longitude of the feature. Additional engineering data regarding the Landsat band and which satellite the image chip was taken from is also included. The accuracy of the point is generally within 20 meters. Figure 2 is an example of a chip file in image format for a path/row in Pennsylvania.

Image correlation is performed using the two dimensional Fast Fourier Transform (2D FFT) computational method to relate ground control points (GCP's) from the CPLBS with the associated locations in each Landsat scene.[8] To initiate the correlation procedure, three points are first identified in the Landsat scene that can also be found on a map. This process is usually done on an interactive display system with the line/sample coordinates found using a trackball cursor. The latitude and longitude of that point is read directly from the map. The three points are used to determine an affine surface that is used as an estimator of where the 2D FFT correlation routine is to search in the image to match a particular GCP. While the affine fit does not give the true location within a pixel (or several pixels), it does provide the search algorithm with a reasonable window in which to search. As good correlations are obtained, the surface is refined so that less searching is required as the algorithm proceeds through the GCP file.

There are several problems associated with using a pre-established ground control point file for image registration. First and foremost, the file has to be built, a large effort that has been expended by NASA and IBM. The file also has to be continuously updated because of changes in the ground scenes and the varying conditions of the imagery. A particularly difficult problem in the Pennsylvania mosaic registration and control effort was trying to correlate the GCP's with Landsat scenes that were acquired over several seasons. The ground reflectance changes that occur from season to season impair the correlation performance. As an example, a stream course feature in a GCP may be highly recognizable in a particular season, but when examining the scene it is being correlated with, the stream may be silted and the

surrounding land cover blends in with the stream creating a low variance, and hence a low information content image. This makes it difficult to correlate all the GCP's selected for that particular path/row. At most, 18 of the 25 GCP's for each path/row were correlated for the Pennsylvania mosaic. One alternative to this problem is to include the GCP's from neighboring paths/rows for the correlation process, however, this was not done due to time constraints.

Successful use of the CPLBS is also dependent upon the geographic locale of the scenes to be planimetrically controlled. In humid continental and humid subtropical climates, atmospheric moisture contributes to haze in the object scene. The deciduous forests, characteristic of these climates, provide a fairly dynamic land cover association posing difficulties in correlating single date GCP's. Experience in constructing Landsat mosaics of the western United States has shown that arid environments produce the most consistent and haze-free imagery, as well as static imagery in terms of overall ground cover. This considerably minimizes problems of poor correlations due to land cover change.

Our experience with digital mosaicking has shown the CPLBS files of ground control points considerably reduce analyst efforts in compiling the ground control point file for the mosaicking process. In addition, the CPLBS files provide a consistent source of LAT/LONG type data whereas 'manually' selected GCP's are subject to numerous errors, due in part to the tedious nature of selecting the points as well as analyst fatigue.

The ground control points correlated with the Landsat scenes used for the mosaic give each scene its position and projection in the global mapping output grid. If each scene was corrected and inserted into the grid with only the GCP's as control, overall planimetric accuracy would be acceptable, but in all likelihood the edges between the neighboring frames would not match perfectly. To remedy this situation, a series of edge matching points are correlated in all overlapping areas of all scenes used. These points are then mapped (controlled) by the GCP's. The net effect of these additional points is to eliminate any side-to-side or top-to-bottom mismatch between scenes.

Information in the overlap area regarding brightness is also obtained and used to radiometrically correct the imagery at the same time that geometry changes are made. Difficulties in matching neighboring scenes radiometrically were experienced during the processing. With haze problems and the varying dates of the imagery, it was possible with existing software to match the brightness (but not variance) of average areas. However, with variance differences not resolved, marked divisions between scenes occur.

The early date mosaic was completed in two stages. Separate control point files and mapping were used for UTM zone 17 and UTM zone 18. The resultant 'halves' of the mosaic for the state were each 6500 lines by 8500 samples. All four Landsat bands were corrected. The Landsat mosaics for each band, and zone, once completed, were segmented into standard map quadrangles. Figure 3 shows the quadrangles within the state. Most quadrangles were one degree of latitude by two degrees of longitude, except for the border quads in the western part of the state. Typical size of an output quadrangle is 3100 lines by 3100 samples. Figure 4 is an example of a 1° x 2° quadrangle while Figure 5 depicts the zone 17 mosaic.



### 3. FOREST/NON-FOREST MOSAIC

In a parallel effort, Goddard Space Flight Center personnel applied multi-spectral classification techniques to the unprocessed Landsat scenes that were used as input for the early date mosaic. One file of data depicting forest and non-forest land cover was derived and sent to JPL to be registered with the mosaic data base. Since the classification was derived from the 'raw' unlogged data, logging was performed using nearest neighbor interpolation to make the nominal geometric adjustments and then geometrically corrected a second time with nearest neighbor interpolation using the control points produced for the early date mosaic. These data were then mosaicked and segmented into the 1 degree by 2 degree quadrangles.

### 4. LATE DATE MOSAIC - POST DEFOLIATION

Requirements for this task stipulated that once the base mosaicking was completed for the entire state, the technology to update the mosaic on a yearly basis be transferred to the State of Pennsylvania. The VICAR/IBIS software system was obtained from COSMIC[6] by the Office of Remote Sensing of Earth Resources (ORSER) at Pennsylvania State University. In early 1982 the system was installed and tested. Additional program modules needed to produce update mosaics were also delivered, installed, and tested. Once the system was running, a test mosaic was attempted with several goals in mind. First, it was necessary to initiate the ORSER staff in the functions and operation of the VICAR system with regard to mosaicking applications. Second, the Penn State computer system was exercised with VICAR to isolate problems peculiar to the facility. Finally, a prototype procedure for actually creating update mosaics had to be generated and an application case performed.

The late date mosaic, as was the early date mosaic, had to be generated in two sections, one section for each UTM zone in the state. In order to ease scheduling difficulties and to provide Penn State ORSER staff with mosaicking experience, a parallel effort was undertaken with the update mosaic for UTM zone 17 being generated at JPL and the update mosaic for UTM zone 18 generated at ORSER.

The Landsat scenes used in the zone 17 update mosaic were, fortunately, in the EDIPS format, easing pre-processing efforts. Table II depicts the scenes used in the update mosaic. Since the second date imagery is registered to the early date mosaic, the resultant products are identical to the original mosaic, except for ground cover changes. The update mosaic's dimensions are the same as the early date mosaic, 6500 lines by 8500 samples, and it is also segmented into the requisite quadrangles.

### 5. ACCURACY

The accuracy of Landsat digital mosaics has been evaluated to some degree by several sources, including Goddard and Purdue University.[9] Edge-to-edge matching is the most visible error in mosaics. Edge errors tend to encourage scrutiny and degrade the aesthetic and planimetric qualities of the final product.

Table II

PATH	ROW	SCENE IDENTIFICATION	LOCATION NAME	DATE
19	31	22311-15214	Titusville	May 24, 1981
19	32	22311-15220	Steubenville	May 24, 1981
18	31	22400-15142	Warren	August 18, 1981
18	32	22400-15144	Pittsburgh	August 18, 1981
17	31	22381-15084	Williamsport	July 30, 1981
17	32	22381-15090	Harrisburg	July 30, 1981

Overall, scene-to-scene mismatch in the Pennsylvania mosaic is minimal. What does exist is difficult to assess primarily because imagery of different dates was used to produce the mosaic. Those few areas that did exhibit some degree of mismatch were on the order of one to three pixels, but only for very short stretches (100 pixels). In addition, mismatch areas generally fell outside the Pennsylvania state border and did not adversely impact the project.

#### 5.1 Planimetric Accuracy

From a cartographic viewpoint, the evaluation of map accuracy represents a difficult procedure. Accuracy is interpreted from map specifications and standards, but several interpretations of the standards is possible depending upon the method used. The gray areas of interpretation must be acknowledged so that the relatively narrow standards are not applied inappropriately, that is, so they do not reflect the intent or spirit of the specifications.

For continuity, the United States National Map Accuracy Standard (NMAS) were applied in a limited way to evaluate the planimetric qualities of the mosaic. These standards are:

For maps of the scale of the scale of 1:20,000 and smaller, not more than 10 percent of the points tested shall be in error greater than 1/50 inch. These limits of accuracy shall apply in all cases to positions of well-defined points only. Well defined points are those that are easily visible such as the following: monuments or markers, such as bench marks, property boundary monuments; intersections of roads, railroads, etc.; Features not identifiable on the ground within close limits are not to be considered as test points within the limits quoted, even though their positions may be scaled closely upon the map. In this class would come timber lines, soil boundaries, vegetation associations, etc.[10]

The root mean square error (RMSE) for identifiable points in a series of 7-1/2 minute quadrangles was calculated. Verification points were located in 19 quads within a 1° x 2° quadrangle in the state. There are over 800 7-1/2-minute quads in Pennsylvania making it expensive to sample each one. For several of these quadrangles, the actual GCP's for the CPLBS were obtained, providing some measure of control. In gathering the data to calculate the RMSE, the goodness of the actual GCP's were examined and found to be excellent per specifications for the CPLBS. Line/sample values for a given point in the mosaic were located 'after the fact' on an interactive display unit with a trackball cursor and then recorded. The calculated position of that point per the UTM mapping projection grid was compared against the located point and the deltas (X,Y) noted. The RMSE was calculated by the following formulae for all points checked:

$$RMS_{LINE(Y)} = \sqrt{\frac{Y_i^2}{n}} ; \quad (1)$$

$$RMS_{SAMPLE(X)} = \sqrt{\frac{X_i^2}{n}} ; \quad (2)$$

$$D = \sqrt{RMS_Y^2 + RMX_X^2} . \quad (3)$$

Results of these calculations are given in Table 3.

Table 3.

	ROOT MEAN SQUARE ERROR (RMSE)	
	PIXELS	METERS
Delta Line	1.13	64.41
Delta Sample	3.49	198.93
Delta D	3.67	209.19

The total number of points used in the verification was 19, one point for each 7-1/2 minute quadrangle. The distribution for these points was narrow; all fell within a 1° x 2° quadrangle. While in the process of the initial verification it was noted that certain areas of the mosaic had geometric stability problems, while others did not. Our efforts were concentrated on the problem areas.

The acceptable error for maps of the 1:250,000 scale class is 127 meters in the X and Y directions. While the line errors are well within this limit, the sample errors are not, and of course neither are the derived D values. These particular errors have been attributed to the Mirror Scan Velocity Profile (MSVP) of the multispectral scanner. Formulas used in the nominal corrections of the data were obtained from the published public record. The formulas are determined by instrument bench tests during system pre-flight checks. It is possible that fatigue and wear in the scanner system caused the MSVP to change, and if so, the correcting formula would change similarly. The MSVP can be compensated for during the mosaicking process but it requires an extremely dense network of GCP's, especially within the peaks and troughs of the profile. Contributing factors that inhibit proper correction are the inability to obtain sufficient correlation of GCP's because of changes in land cover, lack of actual identifiable features, and atmospheric conditions.

## 6. CLOSING COMMENTS

Landsat digital mosaicking is an extremely complex and tedious process because of the nature of the data. If Landsat type multispectral data were available in quantity from a framing type sensor, several problems, particularly those relating to geometry, would be minimized. The reality is that because Landsat data are as plentiful as they are, efforts must be directed to increase their utilization in a wide range of applications. Large regional applications pose particular problems of continuity and data organization whenever the study area exceeds the dimensions of the Landsat framing convention. Mosaicking is one solution to a major part of the problem.

Clearly, differences of opinion relative to 'wants' and 'needs' of accuracy will readily surface. Concurrently, an educational process is also occurring as mosaickers learn more about the 'wants' and 'needs' of the user community, and users learn more about the realities of mapping standards. The ability to locate a specific point in a rural area that lacks valid recognizable points to within 200-300 meters (4-6 pixels) is a vast improvement over non-cartographically based imagery. However, every effort should be made to improve geometric stability and performance of digital imagery such as Landsat mosaics.

## ACKNOWLEDGEMENTS

The authors express their appreciation for contributions of related work as well as support of the following JPL personnel: Richard Blackwell; Nevin Bryant; Tom Logan; Jerry Clark; Boris Gohkman, and Kevin Hussey. The authors also thank Ross Nelson, Lisette Dottavio and Darryl Williams of Goddard Space Flight Center, and Brian Turner, George Baumer, and Fred Luce of the ORSER staff at Pennsylvania State University.

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

## REFERENCES

1. Hoy, M.A., "The Gypsy Moth - Here Again," California Agriculture, July 1982, p. 4-6.
2. Nelson, R. F., Dottavio, C. L. and D. L. Williams, Monitoring Insect Defoliation of Hardwood Forests Using Landsat. Internal paper. Goddard Space Flight Center, 1980.
3. JPL work order #Y677-00-00-03-89.
4. Seidman, J. B., and A. Y. Smith. VICAR Image Processing System - Guide to System Use. JPL Publication 77-37, Rev. 1. National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology, December 1, 1979.
5. Zobrist, A. L. and N. A. Bryant, "Elements of an Image Based Information System." Policy Analysis and Information System, Vol. I, No. 2, January 15, 1978: pp 71-90.
6. COSMIC - Information can be obtained from:  
COSMIC  
Computer Center  
112 Barrow Hall  
University of Georgia  
Athens, Georgia  
Telephone: (404) 542-3265
7. Niblack, W., "The Control Point Library Building System," Photogrammetric Engineering and Remote Sensing. Vol XLVII, No. 12, December 1981, pp 1709-1715.
8. Zobrist, A. L., N. A. Bryant and R. G. McLeod, "Technology for Large Digital Mosaics of Landsat Data," Photogrammetric Engineering and Remote Sensing, Submitted for publication, 1982.
9. Bartolucci, L. A. and T. L. Phillips. Quarterly Progress Report on the Digital Information System for the Oruro Department, Bolivia (ATN/SF-1812-BO), Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, November 1981, pp. 16-30.

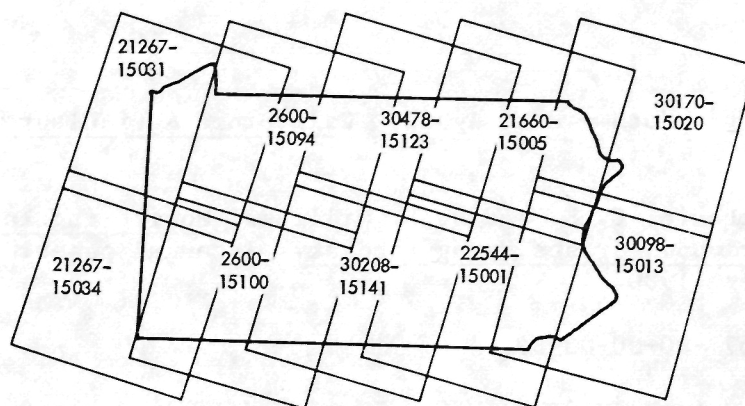


Figure 1. Landsat Frames Footprints - This illustration shows the individual footprints of each scene used for the Pennsylvania Mosaic, both UTM zones.

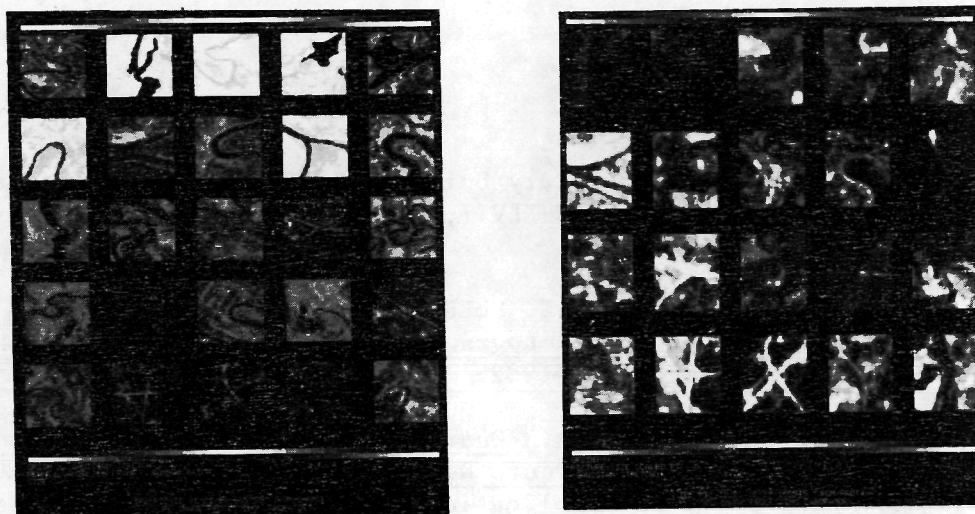


Figure 2. Ground Control Point Images - The GCP's used for controlling the mosaic were obtained from the Control Point Library Building System. The image on the left is the display of the actual CPLBS points, while the display on the right shows the matches in the Landsat scene as a result of the correlation process.

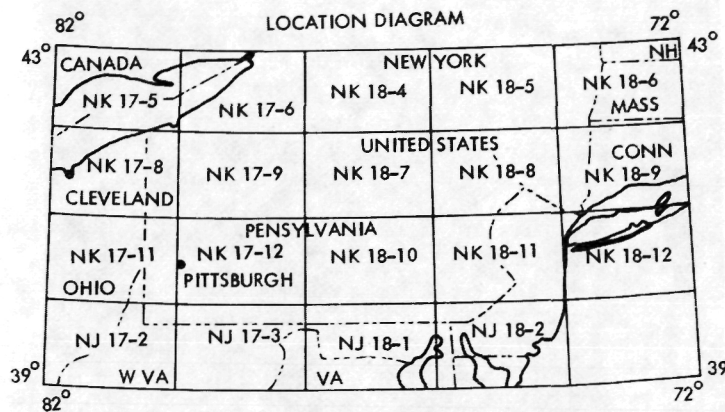


Figure 3. Quadrangles - This location diagram shows the quadrangles used within the State of Pennsylvania.

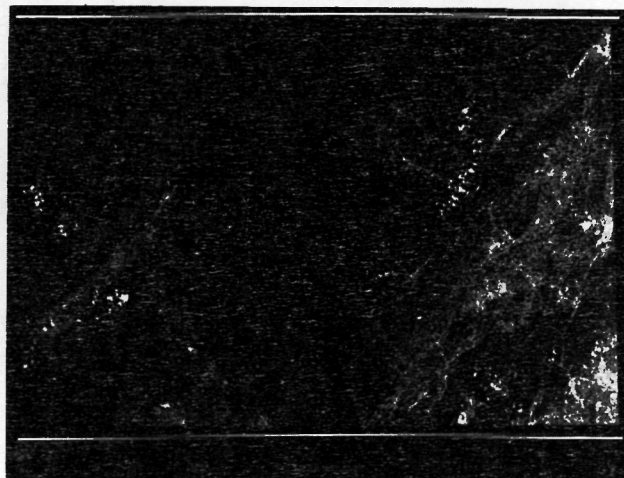


Figure 4. 1 Degree x 2 Degree Quadrangle - Shown here is the Scranton Quadrangle which corresponds to the AMS map series (1:250,000) NK18-8. Landsat Band 5 (red) is displayed.

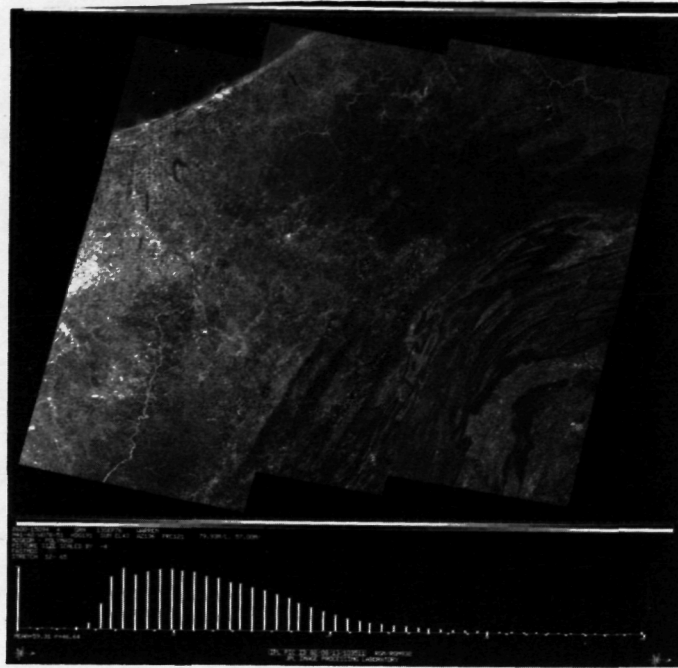


Figure 5. UTM Zone 17 Mosaic - This image depicts the UTM zone 17 Landsat Mosaic for Pennsylvania. Landsat Band 4 (green) is shown here. The image in size is 6500 lines by 8500 samples per line.



**IMAGE ANALYSIS FOR FACILITY SITING:  
A COMPARISON OF LOW- AND HIGH-ALTITUDE  
IMAGE INTERPRETABILITY  
FOR LAND USE/LAND COVER MAPPING**

H.M. Borella  
EG&G, Santa Barbara Operations

J.E. Estes, C.E. Ezra,  
J. Scepan, and L.R. Tinney  
University of California, Santa Barbara

**ABSTRACT**

For two test sites in Pennsylvania the interpretability of commercially acquired low-altitude and existing high-altitude aerial photography are documented in terms of time, costs, and accuracy for Anderson Level II land use/land cover mapping. Information extracted from the imagery is to be used in the evaluation process for siting energy facilities. Land use/land cover maps were drawn at 1:24,000 scale using commercially flown color infrared photography obtained from the United States Geological Surveys' EROS Data Center. Detailed accuracy assessment of the maps generated by manual image analysis was accomplished employing a stratified unaligned adequate class representation. Both "area-weighted" and "by-class" accuracies were documented and field-verified. A discrepancy map was also drawn to illustrate differences in classifications between the two map scales. Results show that the 1:24,000 scale map set was more accurate (99% to 94% area-weighted) than the 1:62,500 scale set, especially when sampled by class (96% to 66%). The 1:24,000 scale maps were also more time-consuming and costly to produce, due mainly to higher image acquisition costs.

**INTRODUCTION**

As technology has advanced to date so too has consumption of energy, an important economic asset in a modern society. The United States and Canada currently consume some one third of the world's total energy production. The need for facilities increases to satisfy new energy demands and shifts in the spatial location of those demands. Mapping of land use/land cover for site potential analysis is an important application of remotely sensed data. The project documented in this paper was designed to analyze the applicability of: 1) aerial photography, 2) appropriate ground supporting data, and 3) site specific scientific literature for use in analysis and

interpretation to meet NRC requirements for facilities siting and transmission line corridor selection (Borella, et. al., 1982). The interpretability of commercially obtained low-altitude and existing high-altitude imagery were compared in terms of time, cost, and accuracy with respect to the preparation of Anderson Level II land use/land cover categories (Anderson, et. al., 1976).

Specific objectives of the research reported herein included:

- Acquisition by commercial means of low-altitude color and color infrared aerial photographs of an area ten miles in radius around two test sites located in Pennsylvania;
- Using these photographs, produce an Anderson Level II land use/land cover map for the ten mile radius area at a scale of 1:24,000;
- Acquisition of the latest available, existing, high-altitude aerial photography through the United States Geological Survey's EROS Data Center and National Cartographic Information Center and map land use/land cover using the Anderson Level II classification scheme;
- Documentation of the time involved in each operation along with the cost associated with each task;
- Conduct of field verification efforts for the two test sites;
- Assessment of the accuracy of the land use/land cover maps generated at each scale;
- Production of a set of map overlays which illustrate the differences between the two scales of maps; and,
- Comparison of the relative time, costs, and accuracies associated with the generation of land use/land cover mapping from the two sets of imagery.

Following a brief discussion of the background of this project, this paper includes sections on: the test sites used for this analysis; the procedures used in data acquisition; the mapping effort; and information on the statistical approach used in the accuracy assessment. Time, cost, and accuracy figures are then compared which contrast the potential of both commercial and existing photography to provide Anderson Level II land use/land cover information. This is followed by a brief section which includes the conclusions arising from findings of this research effort.

## BACKGROUND

The National Environmental Protection Act (NEPA) has been interpreted to require terrestrial and aquatic resource inventories and descriptions of preferred facility sites and transmission corridors before an assessment can be completed and a construction permit issued. NRC Regulatory Guides 4.2 and 4.7 outline information needs and siting considerations in a general way, but leave considerable area for interpretation of specific needs to applicants and their consultants. Because of applicants' uncertainty as to specific detailed information requirements, their response over a period of years has been to generally increase the volume of information submitted in successive environmental reports in the hope of gaining comprehensive coverage. The effect has been and is now to saturate the assessment process with information leading to excessive staff time for review and in some cases to an unintentional obfuscation of issues rather than clarification.

The Council of Environmental Quality (CEQ) has cautioned agencies on this problem in the preparation of Environmental Impact Statements (EIS), and has recommended that these documents should be limited to the essential information needed for rational decision-making. Following the same reasoning, it is believed that the continual growth of environmental reports should be limited by better specification of information requirements or by formats which would satisfy regulatory staff needs for assessment and decision-making, even though the reports are reduced in volume.

Unfortunately, specification of detailed information needs on a point-by-point basis has proven to be a relatively intractable problem because of the site-specific nature of environmental assessment. It has usually been necessary to trade off detailed instructions for general guidance which is applicable on a nation-wide basis. An alternative to detailed descriptive guidance is to specify an informational format which would show certain relevant details about a site. Remote sensing is such a format.

Adoption of remote sensing techniques for regulatory environmental guidance may have the advantage to NRC of enabling agency personnel to specify comprehensive information gathering techniques which are not site-specific but which would, in all probability, yield a substantial portion of the information needed for licensing assessment on any site likely to be considered.

Although development of remote sensing techniques for facilities siting appears desirable, uncertainties exist with respect to the potential of adapting them to regulatory needs. The scope of this work deals specifically with aerial photography, the main uncertainties of which at present relate to the following:

1. The extent to which existing regulatory siting guidance can be met using these methods.
2. Fine tuning of the interplay between aerial photography and ground truthing needed to meet licensing requirements.
3. Quantification of presumed cost advantages of these methods.
4. Relative information return from the technology.

Image acquisition (commercial low-altitude and existing) coupled with field visits were carried out to determine the information return from remote sensing technology in relation to selected regulatory siting requirements, namely land use/land cover. The results presented herein should provide NRC a documentary basis for evaluating these techniques for acquiring information relative to resource evaluation for inclusion in environmental reports and for revising existing guidance for making environmental surveys.

### **TEST SITES**

The two circular test sites with a radius of 10 miles selected for acquisition, analysis, and comparison of the commercially acquired and existing high- altitude, remotely sensed data are located in east central Pennsylvania (see Figure 1). The northernmost site is centered on the Susquehanna power plant site near the town of Berwick, Pennsylvania, Latitude 41 5' 30"N Longitude 76 8' 0"W.

The Berwick site is located in a folded ridge and valley section of the Appalachian Mountain System. This area land cover is predominantly forest with heavy strip-mining activities. Both urban (nearly 7%) and agricultural activities (about 21%) are less dominant in areal extent here than at the Lancaster site. Agriculture areas are mostly in corn and pasture, while both active and abandoned strip mines attest to coal mining in the area.

The second site is Lancaster and its environs, which is located

in south-central Pennsylvania and centered at latitude 40 2' 15"N; longitude 76 18' 20"W (Figure 1). Land use/land cover in the Lancaster site exhibits more of a mixture of cover types. The dominant land use patterns are related to agriculture and related activities (approximately 60% of the total area) and to large urban areas (nearly 22% of the total area). Dominant agricultural activities revolve around corn, oats, hay, alfalfa, tobacco, and some truck crops.

### DATA ACQUISITION

Commercial low-altitude (1:24,000) color and color infrared aerial photography was flown over the two test sites on 25 September 1981 by Photo Science, Inc. (PSI), based in Gaithersburg, Maryland.

It was agreed that mapping land use/land cover in Pennsylvania would best be accomplished after full leaf maturity in the spring and before leaves start to fall in autumn: this image acquisition mission was coordinated by EG&G and University of California, Santa Barbara (UCSB) personnel. EG&G/UCSB developed detailed specifications for the image acquisition mission and coordinated the contract with PSI through Pacific Western Aerial Surveys of Santa Barbara, California. (In the commercial aerial image acquisition field, it is not unusual for individuals needing imagery of a distant area to work through a local firm that in turn contracts with a firm near the area of interest to actually fly the image acquisition mission.)

The camera systems flown by PSI were Wild Heerbrugg RC8's, which exposed color and color infrared films simultaneously. Flight lines followed by PSI were jointly planned by personnel from Pacific Western Aerial Surveys and UCSB, such that sufficient overlap (60%) for stereoscopic coverage was obtained. Sidlap was designated at 20% to ensure that no gaps or misses would occur through lack of coverage from line-to-line or through crab or yaw of the aircraft in flight.

After PSI flew the image acquisition mission, the film was sent to Dayton, Ohio for processing, returned to PSI for quality assurance and assessment, and shipped to Pacific Western Aerial Surveys in Santa Barbara, which received the 1:24,000 scale color and color infrared aerial imagery in late October. These materials arrived in the format of 9x9-inch color infrared (CIR) positive transparencies and color negatives from which positive prints were produced.

When received, the processed low-altitude aerial imagery was

thoroughly checked with respect to mission specifications in the contract. Specifically, the film quality was assessed, degree of vignetting examined, color balance and processing checked, actual flight lines plotted against the flight-line maps provided, photograph centers plotted, overlap and sidelap computed, and correctness of scale verified. Though the color infrared (infrared ektachrome) imagery was of totally acceptable quality, the color color film was acceptable but not of high quality; basically the latter appeared to be overexposed and had a somewhat "washed out" appearance. Although bothersome, the degree of loss of contrast was not sufficient to reject the imagery (part of the UCSB Central Library has computer links to data centers and imagery coverage catalogues). All other characteristics and parameters of the imagery flown by PSI met contract specifications and the data were accepted. Figure 2 is an example of the color infrared aerial photography acquired for this project.

High-altitude aerial photography has been taken by various Federal Government agencies for research and applications-oriented purposes. Information concerning this coverage can be obtained from a number of Federal data banks (e.g., the U.S. Department of the Interior's United States Geological Survey, EROS Data Center, Sioux Falls, South Dakota). A thorough search of the coverage acquired by all Federal Agencies and other sources was conducted through the UCSB Map and Imagery Collections Library (part of the UCSB Central Library has computer links to data centers and imagery coverage catalogues) to determine the availability of post-1974, existing, high-altitude coverage for the two test sites. This imagery search yielded the following results:

1. Black and white high-altitude imagery flown by the United States Geological Survey (USGS) on 10/12/76 at a scale of 1:78,000 for the 10-mile radius around both test sites. This imagery was obtained from EROS Data Center in early November 1981.
2. Color infrared high-altitude imagery flown by the National Aeronautics and Space Administration (NASA) dated 7/21/74 data scale of 1:126,000 of the Berwick site appeared to be available, and this imagery was also obtained from EROS in early November 1981. Examination of the example (Figure 3) shows early construction activity at the Susquehanna power plant. This imagery was judged to be of very good quality by the analysts.
3. Color infrared high-altitude imagery for the Lancaster site flown by NASA on 2/5/74 at a scale of 1:128,000

also appeared to be available. This imagery was also ordered, but only partial coverage of the Lancaster site was available.

Although statements regarding image quality, extent of coverage, degree of cover, etc., are included in the catalog material concerning the high-altitude photography, once this imagery was received it was thoroughly examined for its potential to meet project requirements. Color balance, ground resolved distances (GRD), acutance, scale, overlap and sidelap were all checked to determine not only the degree of site coverage but also the quality of coverage as related to Anderson Level II land use/land cover mapping requirements. Based on evaluation of these data, it was found that:

1. While the USGS black and white image coverage was complete for the two sites, the graininess of the film and its poor contrast made it of very poor quality for land cover mapping. That is, the Principal Investigator and image analysts involved in this project determined that acceptable Anderson Level II land use/land cover mapping accuracies could not be achieved using these data.
2. The NASA color infrared imagery coverage of the Berwick site was indeed complete and, as previously indicated, the analysts deemed the quality of this imagery as appropriate for Anderson Level II land use/land cover mapping.
3. Complete coverage of the Lancaster site was found not to be available, and the poor color balance of that portion of imagery which did cover the site rendered the data unacceptable for detailed land use/land cover mapping.

#### COLLATERAL DATA

No two areas of this country are exactly alike physically, socially, or culturally. All areas have a certain degree of uniqueness in their land use/land cover patterns. It is axiomatic that the more familiar an image interpreter is with the region and/or phenomena he/she is asked to analyze, the more accurate the analysis will be. As such, on-site field verification visits are often important in any image analysis project, particularly in those which the analysts interpret data acquired in areas about which they have limited knowledge. As the image analysts in this project could not visit the sites prior to the data analysis phase of this project they were, in

essence dependent upon collateral data for site-specific background information. These collateral data included:

1. U.S. Geological Survey generated Land Use and Land Cover (LUDA) 1:250,000 maps for the two test sites. The Lancaster test site is located on the 1972 Harrisburg quadrangle, and the Berwick (Susquehanna) site is found on the 1972 Williamsport quadrangle.
2. U.S. Geological Survey 1:24,000 scale (7.5') and 1:62,500 scale (15') topographic quadrangles.
3. Vegetation maps and related articles.

### **MAPPING**

Areas within a radius of 10 miles of the Lancaster and Berwick test sites were mapped at Anderson Level II using the commercially acquired color infrared transparencies. (Figure 4 shows indices of the 1:24,000 scale 7.5' USGS topographic quadrangles relevant to the Berwick study area.)

Land use/land cover mapping was accomplished using manual photo interpretation procedures. The color infrared transparency imagery was solely employed in the interpretation; these data were deemed of superior quality for Anderson Level II mapping purposes compared to the color print data. To aid in location and area referencing during the land use/land cover mapping, mylar overlays with roads and stable features copied from 1:24,000 USGS topographic quadrangle maps were used as a mapping base; i.e., the mylar overlays to be used directly with the aerial image were annotated with roads and other stable features to make the interpreters' information transfer task more efficient.

Trained imaged analysts interpreted the imagery and transferred the interpreted land use/land cover classes to the base map thereby producing pencil-line working copy maps. The interpreters who analyzed the low-altitude imagery at one site did not work on the high-altitude imagery of the same site, which eliminated potential bias which might have been generated by interpreters becoming familiar with a site through experience gained at one scale, and transferring this in their analysis of another scale.

### **MAPPING FROM EXISTING DATA**



Due to lack of comprehensive coverage of the total area of both test sites by either of the two sets of available color infrared imagery (making statistically significant comparisons with the low-altitude data more tenuous), a decision was made to use the black-and-white (B&W) data to map both sites.

The data were received in a B&W positive transparency format (1:78,000 scale). Pacific Western Aerial Surveys made and printed inter-negatives to the scale of 1:62,500. Overlays of 15' topographic quadrangle maps were generated and used as the base maps. Interpretations were made from the original B&W positive transparencies. The prints were used for locating and mapping features.

After considerable analysis it was determined that Anderson Level II criteria could not be met employing these data. As such, Anderson Level I criteria were used. Problems have mainly related to image quality, interpretability, consistency of category identification, and labeling, which rendered uniform applications of Level II criteria difficult, if not impossible; that is, it was determined, to the author's satisfaction, that Anderson accuracy criteria could not be met at Level II.

A Level II classification was accomplished for the Berwick site using the color infrared high altitude imagery, which was examined, evaluated, and deemed of acceptable quality to meet Anderson Level II accuracy requirements. Analysis of these data was accomplished employing techniques similar to those described above.

Once pencil line "working copy" maps had been employed in the field verification and accuracy assessment process, final "archive copy" maps were generated from them. Land use/land cover polygons were transferred from the pencil line working copy maps to final archive maps. It should be emphasized that no corrections to the working copies were made as a result of field verification.

An example of a final map product at (1:24,000 scale) is shown in Figure 5.

#### **MAPPING PROBLEMS**

Local relief distortion inherent in the low-altitude aerial photography acquired from PSI sometimes caused slight mislocation of smaller features on the map. This localized "scale error" required that the image analysts essentially manually "rubber sheet" the base map overlays to the imagery in order to accurately map photo-derived information to their

correct planimetric positions.

Up-to-date 1:62,500 scale USGS topographic maps for the two Pennsylvania sites were not available. The most recent map was updated in 1955; the earliest in 1981.

The 1:62,500 scale black-and-white (B&W) imagery was of poor quality. As previously stated, this B&W imagery exhibited very low contrast, which reduces texture in some features making them less interpretable, and also tends to blur the edges of features. This loss of acuteness makes object identification more difficult, particularly when dealing with small area polygons at Anderson Level II.

Lack of available color imagery increased difficulty of mapping vegetated versus non-vegetated areas in transitional regions. Scale-relief distortion was also found in the existing imagery.

Given the low contrast, the scale of this imagery was generally too small to permit Level II mapping consistency throughout the study area and to meet minimum mapping unit standards.

Finally, there was also some difficulty caused by having to map with opaque prints (with back-lighting) as opposed to transparencies (although all interpretation was done by viewing the original B&W positive transparencies and transferring information to the mylar overlays).

Complete Lancaster site coverage was not available, and the existing coverage (eastern portion of study area) was determined to be of very poor quality (poor color saturation, contrast, and extensive vignetting).

Lack of consecutive coverage of both sites with acceptable quality color infrared imagery also makes statistical comparison with maps generated from 1:24,000 scale data less than satisfactory. Therefore, Anderson Level II accuracies achieved in this project can only be compared for the Berwick site.

Along with the 1:24,000 scale and 1:62,500 scale Level II classification maps, an additional map was created for the Berwick site. This overlay was produced to locate, identify, and analyze classification discrepancies between the land cover maps at the two mapping scales.

The comparative difference overlay was produced as follows: each of the twelve 1:24,000 scale archive classification maps were photographically reduced onto a clear film medium at 1:62,500 scale. The separate reduced maps were then overlain on top of the archive copies of the 1:62,500 scale

classification maps. Areas exhibiting differing classifications between the 1:24,000 and 1:62,500 scales were then traced onto a separate sheet of frosted mylar. A minimum mapping unit of 0.5 inch (0.25 cm) was used, and special care was taken to insure accurate registration of the two maps (see Figure 6).

The final map product of this effort is a single, frosted mylar sheet showing the areas which were found to exhibit classification discrepancies between the 1:24,000 scale Anderson Level II mapped classes and the 1:62,500 Anderson Level II mapped classes for the Berwick site.

An examination of this 'difference' map shows that the Anderson Level II classes which appear most frequently are residential, cropland and pasture, other agricultural land, and deciduous forest. Reviewing each class in turn, it is possible to speculate why these differences might occur.

That the mapping was done by different photo interpreters using imagery flown almost six years apart is the most obvious cause of differences in classifications between the two map series. Although errors caused by some difficult-to-document land use changes are bound to come into the spatial errors on the discrepancy map, overall accuracies for both scale maps were quite high within this project. This discussion and documented errors are not statistically significant in the total context of the mapping effort.

Class Residential, a category often found surrounding other types of land uses (i.e., commercial and mixed urban classes), is also scattered throughout the agricultural and forested areas of the Berwick study area in the form of a single unit residences (often with smaller detached structures). In addition, identifying residential land uses and structures sometimes requires considerable use of "collateral data".

Category Cropland and Pasture would seem to be a distinct and fairly easy classification to map, but in fact, the Anderson scheme leaves room for subjective interpretation. Also, some of the agricultural practices (e.g., small field, diverse crop farming, etc.) peculiar to this study area can make identification of cropland and pastureland difficult.

Because Class Other Agricultural Land is a very broadly defined category in the Anderson scheme, the photo interpreter must make certain basic decisions at the beginning and follow them throughout the mapping effort in a systematic way. Elements of this category (perhaps more than any other) can be placed legitimately in other classes within the Anderson scheme.

Different image analysts, unless they have worked together or have decided on specific guidelines prior to the interpretation task, can often have difficulties in consistently labeling this category.

The Deciduous Forest Category, the most widespread land cover type present in the Berwick site, is a constant target for change as forested areas are given over to agricultural and urban/suburban land uses.

### **STATISTICAL PROCEDURES FOR ACCURACY ASSESSMENT**

The statistical basis of the accuracy assessment procedures are basically those presented in detail in "Sampling for Thematic Map Accuracy Testing," an article by Rosenfeld, et. al., (1982), in the January issue of Photogrammetric Engineering and Remote Sensing, which describes a stratified systematic unaligned sampling technique. Sample points taken from the map as a whole, with additional random samples for under-represented thematic categories, are used to estimate the thematic accuracy of all mapped categories.

Prior to selection of sample points to be verified, the minimum sample size needed to validate the accuracy for each category within specified confidence limits was estimated using a cumulative binomial distribution. The binomial distribution is proper in this case as verified points can be either "correct" or "incorrect". Anderson, et. al., (1976) state, "the minimum level of interpretation accuracy in the identification of land use and land cover categories from remote sensor data should be at least 85 percent".

After preliminary evaluation of the commercial color infrared photography and the existing high-altitude color infrared imagery of the Berwick site and a review of the Anderson Level II classification scheme, all interpreters felt confident that the 85% accuracy level from these data could be attained. Based on a detailed analysis, it was also determined that Anderson Level II accuracy criteria could not be met using the existing B&W high-altitude photography. However, Level I Criteria could be met. Anderson Level I maps of both sites were generated, but do not relate to the discussion of map accuracy presented herein.

Using the binomial distribution and imposing a 95% confidence requirement as described by Rosenfeld, et. al., (1982), it was calculated that a minimum sample size of 19 points per thematic category per site was needed to verify Anderson Level II classifications at 85% accuracy, with an allowable error of 10%.

Once the desired minimum sample size per category was determined, the sample selection procedure was implemented as follows:

1. Each site was stratified using a 5-km grid network based upon Universal Transverse-Mercator (UTM) coordinates. These 5-km square "strata" provided the basis for subsequent sampling.
2. A systematic sample grid overlay was made for each scale of strata used (1:62,500 and 1:24,000 scale maps were used). The overlay was partitioned into a 500-m sample grid covering the 5-km strata, thus resulting in a 10 by 10 sample grid of 100 sample points per strata coded by number as shown in Figure 7.
3. A computer program was used to randomly order the 100 potential sample points within each strata. A separate random sample was provided for each strata thereby resulting in an unaligned sample design since not all samples were used.

Two sets of points were generated for each map to provide both area-weighted and category-specific estimates of accuracy. (An area-weighted accuracy assessment tests accuracy of the map as a whole. This technique yields an overall accuracy figure for the entire map without regard format accuracies within individual classes. A "by class" sampling technique provides accuracy figures for each thematic category within the map.) An initial set of points was selected for verification by taking the first five random points within each 5-km strata, and marking their locations on the working map copies (see Figure 9) in the generation of approximately 175 points per site to be used in an area-weighted accuracy estimate. After tabulating the land use/land cover category of each point, a second set was generated in an iterative manner thereby subsequent groups of five points per strata were examined for all strata, and points for any under-represented categories were added to the list to be verified and flagged as such on the working copy map overlays. This dual approach provides both an accurate area-weighted sample and an efficient set of additional points needed for category-specific accuracy assessments. In some instances, however, it was not possible to adequately represent some classes due to their relative rarity, even after exhausting all 100 potential sample points per strata.

Following Rosenfeld's procedure, initial verification of

mapping accuracy was done by an expert photointerpreter who hadn't participated in the original mapping effort. Any point he was unsure of was designated for subsequent field verification; obvious agreements and outright errors were directly photo-verified and tabulated. This procedure is, of course, constrained by the level of expertise of the expert photo interpreter. Subsequent field verification efforts basically validated this approach, but the lack of local area familiarity did result in one important category verification error involving Deciduous and Mixed Forest classes in the Berwick site. The impact of this problem, which was limited to the one site, is discussed next.

In the field, it was estimated that approximately 10-15% of the Berwick site is actually Mixed Forest and not Deciduous Forest as mapped and photo-verified. Since neither the photointerpreters nor expert verifier were very familiar with the area, this is somewhat understandable. With only single dates of imagery available at scales where the resolution of individual tree crowns is difficult, such errors can and do occur. This Mixed Forest category is a problem to USGS land use/land cover mappers as well. In this study, interpreters made logical conservative decisions based on their experiences. Had they been provided with both summer and winter images of these areas, classification errors would have been reduced.

Field verification involved approximately 20 specified samples per site with an additional 10 "correct" points added to test the expert photo-interpreter's verification accuracy. All "correct" points did agree in the field.

Accuracy results are presented in Tables 1 through 6. In each instance an area-weighted accuracy estimate table is provided first, followed by per class accuracy estimates. The range of numbers for each table of estimate represents the range corresponding to a 95% confidence level. Given the results found for the specified sample size, we can expect the true estimate to fall within this range 95 out of a 100 times.

Because the Deciduous and Mixed Forest Classes in the Berwick site are not adequately distinguishable using the available imagery, we can only estimate the mix of those classes. Examining the USGS LUDA maps we find that 69.3% of the forested area is mapped as Deciduous Forest and 30.7% is classed as Mixed Forest. Assuming this relationship is correct, we can estimate that 19.4% of our 1:24,000 scale map is in error where Mixed Forest areas have been mapped as Deciduous Forest.

#### **TIME AND COST FACTORS**

Table 7 is a summary of time involved in creating base map overlays and photo-interpretation. Time factors were broadly similar between sites, with slightly more time involved for the Lancaster site than for the Berwick site. Analysts considered this increased time to be due primarily to the greater amount of urban and agricultural land use in the Lancaster site. Interpretation and mapping time were generally related to the amount of map area involved; the increased time required to complete the 1:24,000 scale maps, as compared to the 1:62,500 scale maps, is basically the same factor (about 2.5x) as the map area differences involved.

The time and cost factors are relevant only in that they give an estimate of the overlay and interpretation time by the analysts. It does not cover all the hours of technological supervision provided by the Principal Investigator. Table 7 basically summarizes costs for a first effort, not idealized because of lack of area visitation, knowledge of the area, etc.

Subsequent activities could be more economical, but may not be because the sites of interest may be unique in their topographic, demographic, and sociological parameters.

While it appears that the cost of the 1:24,000 imagery for the two sites has a cost/site of approximately \$6300, this reflects only the contract for the flying, some planning, acceptance, etc. It does not include the many overview hours by the Principal Investigator and others in a cooperative effort in flight line planning, outlining photographic specifications and then, after the film is returned, the overlap, sidelap, altitude, etc., checking and acceptance. The cost of approximately \$6300 should be viewed as the image acquisition cost only.

## CONCLUSIONS

Based on an overall assessment of time, costs, and accuracies in the study of two sites in Pennsylvania, it is believed that commercially acquired 1:24,000 scale low-altitude aerial photography is preferable for use in Anderson Level II mapping projects of this type. Although more expensive in terms of cost (especially in respect to actual data acquisition), the 1:24,000 scale imagery is definitely more interpretable and accurate, especially if category accuracy is important. For the siting task of interest to NRC, the higher accuracy of the low-altitude image based mapping may also have important legal implications. The uncertainty of availability of current, high-quality, high-altitude photography of a chosen area is another reason for choosing the 1:24,000 scale imagery.

In addition, using available high-altitude imagery as a base may necessitate mapping at scales smaller than appropriate for the given siting tasks. That is, the use of smaller scales in mapping implies using a larger minimum mapping unit and a subsequent loss of information in some categories throughout the map set.

#### **ACKNOWLEDGEMENT**

Work on this paper was conducted under Nuclear Regulatory Commission Contract No. A1251-2 to Energy Measurements Group Santa Barbara Operations EG&G. The authors of this paper wish to acknowledge the interest and assistance of Doctors Philip Reed and Jerry LaRoche of NRC in this project. Their suggestions during data-acquisition/analysis and report writing were most helpful.



# MAPPING TEST SITE LOCATIONS, PENNSYLVANIA

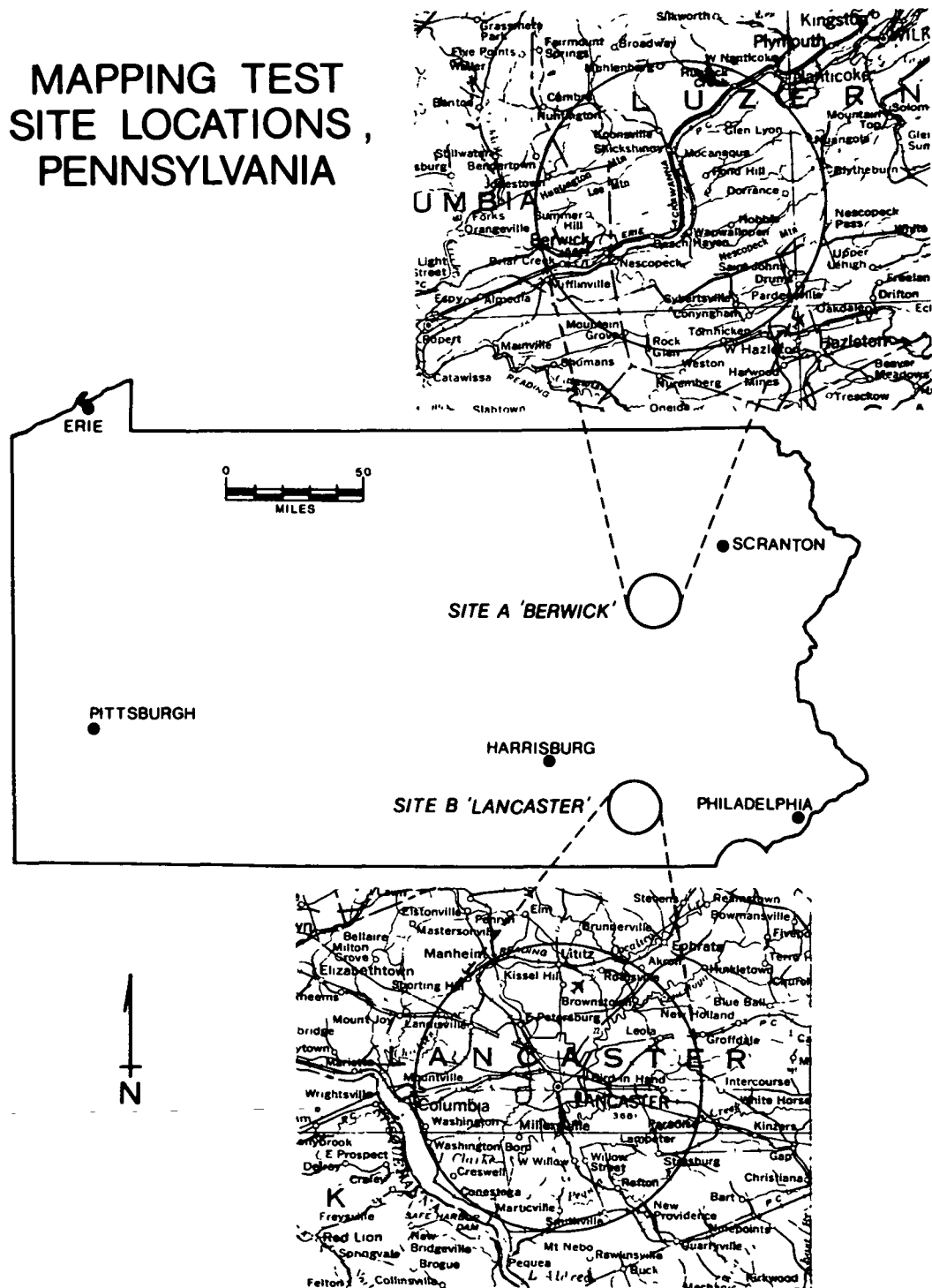


Figure 1. Diagram of mapping test site locations, Pennsylvania.

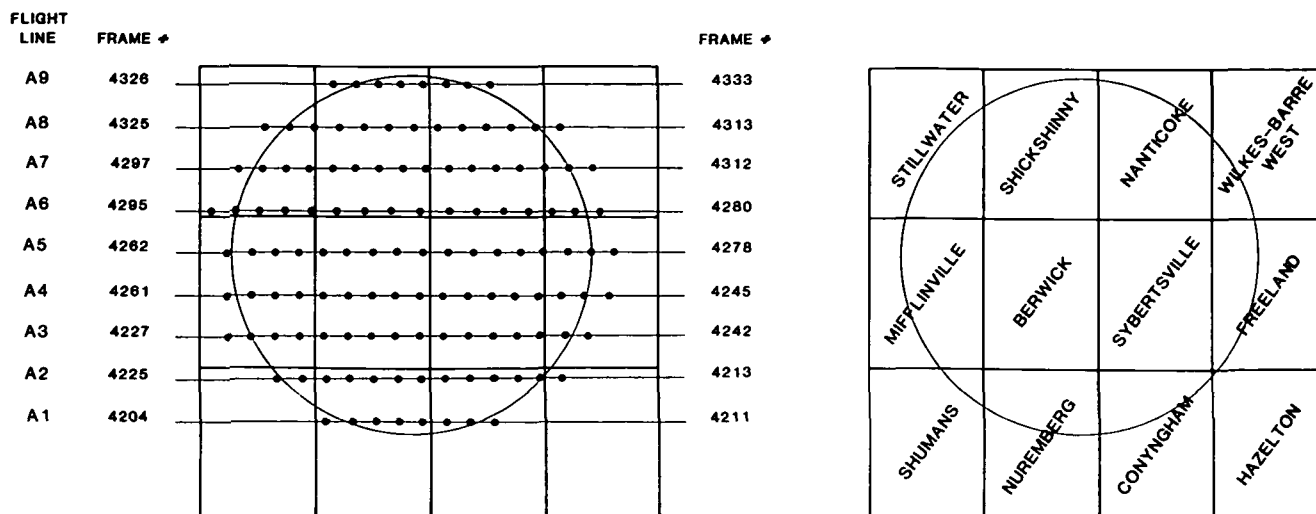


Figure 2. Sample of a portion of color infrared photography acquired 9/25/81, scale 1:24,000, Frame 4270, Area A, showing the extremely high quality and fine detail of CIR photography acquired for mapping effort. Susquehanna power station and surrounding area are shown.



Figure 3. Example of a portion of 1:126,000 scale color infrared photography (7/21/74) acquired for project (also from USGS EROS).

# CIR PHOTOGRAPHY AND 7 1/2 MIN QUAD INDEX - SITE A (BERWICK)



FRAME NUMBERS REFER TO FIRST AND LAST FRAMES IN EACH FLIGHT LINE -  
 1:24,000 SCALE COLOR INFRA-RED PHOTOGRAPHY FLOWN SEPT 25, 1981

Figure 4. Flight line/topo map locator: Area A



Figure 5. A portion of 1:24,000 scale 'archive copy' finished map. Anderson Level II classification. Figure shown is not at 1:24,000 scale.

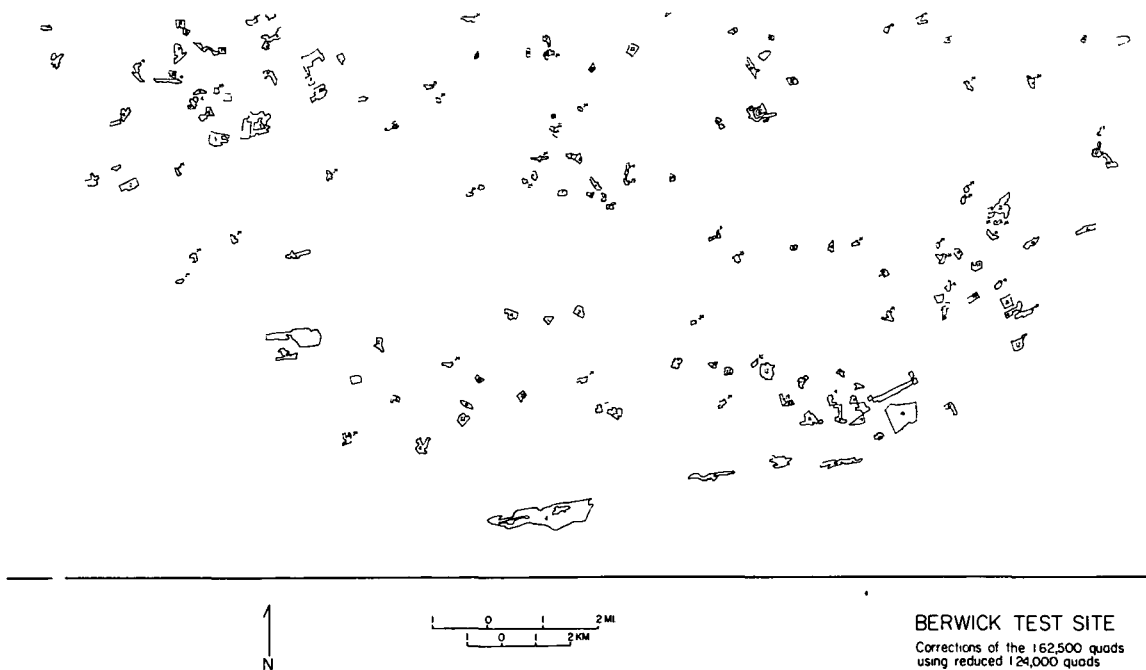


Figure 6. Portion of Comparative Difference Map.

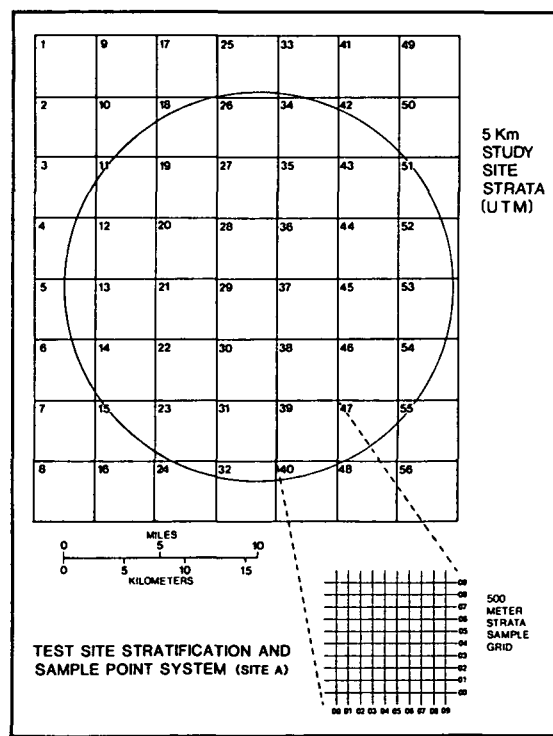


Figure 7. Systematic sample grid overlay employed during the accuracy assessment phase of the project.

		MAPPED CLASS																			
		11	12	13	14	15	16	17	21	22	23	24	41	51	52	53	61	62	72	75	76
VERIFIED CLASS	11	7							1												
	12		1																		
	13			0																	
	14				3																
	15					0															
	16						0														
	17							1													
	21								35												
	22									0											
	23										0										
	24											1									
	41												109								
	51														2						
	52															2					
	53																0				
	61																	1	0		
62																			0		
72																				0	
75																				1	
76																					8

OVERALL CLASSIFICATION ACCURACY 170/172 = 99% \*

95% CONFIDENCE INTERVAL 96 - 100% \*

\*does not include correction for coniferous or mixed forest lands  
problem discussed in text

Table 1. Berwick low altitude mapping (1:24,000 scale) accuracy. Area weighted sampling. USGS Anderson Level II classification.

	MAPPED CLASS																CLASS ACCURACY	95% CONFIDENCE INTERVAL					
	11	12	13	14	15	16	17	21	22	23	24	41	51	52	53	61			62	72	75	76	
VERIFIED CLASS	11	35																			100%	90-100%	
	12		22																		100%	84-100%	
	13			20																	100%	83-100%	
	14				19																100%	83-100%	
	15					0															*	*	
	16						22														100%	84-100%	
	17							22				1									96%	78-100%	
	21	1						1	33	1										2	87%	70-95%	
	22									23											100%	85-100%	
	23										14										100%	77-100%	
	24										1	1	23								92%	73-99%	
	41												115								1	99%	95-100%
	51													21	1							96%	77-100%
	52													1	23	1						92%	74-99%
	53															27						100%	87-100%
	61																1	5				83%	*
62																	1	20		1	91%	71-99%	
72																		0			*	*	
75																			16		100%	79-100%	
76											1			1					1	40	93%	80-99%	
																						96%** AVERAGE CLASS ACCURACY	

96%\*\* AVERAGE CLASS ACCURACY

\*inadequate sample for meaningful assessment

\*\*does not include correction for coniferous or mixed forest lands  
problem discussed in text

Table 2. Berwick low altitude mapping (1:24,000 scale) accuracy. Sampling by classes. USGS Anderson Level II classification.

	MAPPED CLASS																				
	11	12	13	14	15	16	17	21	22	23	24	41	51	52	53	61	62	72	75	76	
11	7																				
12	1	0																			
13			0																		
14				2																	
15					0																
16						0															
17							1														
21	1							35													
22									0												
23										0											
24											0										
41								2				109								2	
51													1								
52														2							
53															0						
61												1				0					
62																	0				
72																		0			
75								3				1							1		
76																				4	

OVERALL CLASSIFICATION ACCURACY 162/173 = 94% \*

95% CONFIDENCE INTERVAL 89 - 96% \*

\* does not include correction for coniferous or mixed forest lands problem discussed in text

Table 3. Berwick high altitude mapping (1:62,500 scale) accuracy. Area weighted sampling. USGS Anderson Level II classification.

	MAPPED CLASS																					
	11	12	13	14	15	16	17	21	22	23	24	41	51	52	53	61	62	72	75	76		
11	50							2				1									94%	84-99%
12	6	4	1			1		2													29%	8-57%
13	4	1	6			2		1													43%	15-69%
14				18																	100%	82-100%
15					0																*	*
16	5					6		1													50%	21-79%
17	7					1	8					1									44%	21-70%
21	1							94													99%	95-100%
22	1							1	4										1		57%	*
23	1							1		0											*	*
24	3										0										*	*
41								2				131							2		97%	92-99%
51												1	17								94%	73-100%
52								1				1	11	5							58%	35-80%
53								1						19							95%	75-100%
61								1				2				0			1		*	*
62								1					5				3		4		23%	4-50%
72																		0			*	*
75								4				1		1				10	2		56%	40-91%
76	3	1						11				3						1	16		46%	29-65%

66%\*\* AVERAGE CLASS ACCURACY

\* inadequate sample for meaningful assessment

\*\* does not include correction for coniferous or mixed forest lands problem as discussed in text

Table 4. Berwick high altitude mapping (1:62,500 scale) accuracy. Sampling by classes. USGS Anderson Level II classification.

		MAPPED CLASS																									
		11	12	13	14	15	16	17	21	22	23	24	41	51	52	53	61	62	72	75	76						
VERIFIED CLASS	11	29																									
	12		5																								
	13			1																							
	14				1																						
	15					0																					
	16						1																				
	17							1																			
	21								90		1	1															
	22									2																	
	23										7																
	24											2															
	41												22														
	51														9												
	52															0											
	53																0										
	61																	0									
62																		0									
72																			0								
75																				0							
76																					0				3		

OVERALL CLASSIFICATION ACCURACY 173/176 = 98%  
95% CONFIDENCE INTERVAL 95 - 99%

Table 5. Lancaster low altitude mapping (1:24,000 scale) accuracy. Area weighted sampling. Anderson Level II classification.

	MAPPED CLASS																				CLASS ACCURACY	95% CONFIDENCE INTERVAL	
	11	12	13	14	15	16	17	21	22	23	24	41	51	52	53	61	62	72	75	76			
VERIFIED CLASS	11	51																				100%	93-100%
	12		24																			100%	85-100%
	13			1	20																	95%	76-100%
	14					20																100%	83-100%
	15						1															*	*
	16								17													100%	81-100%
	17			1	1					14												88%	61-98%
	21			1			1			1	101							1		1	1	94%	88-98%
	22											20										100%	83-100%
	23												26									96%	86-100%
	24													25								100%	86-100%
	41														40							100%	91-100%
	51															26						100%	86-100%
	52																1					*	*
	53																	20				100%	83-100%
61																		0			*	*	
62																			6		100%	*	
72																				11	*	*	
75																			8		100%	*	
76																			2	19	86%	63-98%	

\*inadequate sample for meaningful assessment

97% AVERAGE CLASS ACCURACY

Table 6. Lancaster low altitude mapping (1:24,000 scale) accuracy. Sampling by classes. USGS Anderson Level II classification.

TIME AND COST FACTOR COMPARISONS FOR LEVEL II MAPPING

	1:24,000 SCALE				1:62,500 SCALE			
	OVERLAYS		INTERPRETATION		OVERLAYS		INTERPRETATION	
	TIME (HRS)	COST (\$)	TIME	COST	TIME	COST	TIME	COST
BERWICK	16	384.00	70 1/2	1692.00	6	144.00	20 (I) 31 (II)	480.00 744.00
LANCASTER	25 1/4	606.00	78	1872.00	7 1/2	180.00	24	576.00

(I) ANDERSON LEVEL I MAPPING

(II) ANDERSON LEVEL II MAPPING

FIGURES REPRESENT COSTS OF WORK PERFORMED AT  
UC SANTA BARBARA BY REMOTE SENSING UNIT STAFF

Table 7. Time and cost factors comparisons for Level II mapping.



## BIBLIOGRAPHY

- Anderson, James R., Ernest E. Hardy, John T. Roach, and Richard E. Witmer, "A Land Use and Land Cover Classification System for Use With Remote Sensor Data," U.S. Geological Survey Professional Paper 964, 28 pages (1976).
- Avery, Thomas E., Interpretation of Aerial Photography, 3rd Edition, Burgess, Minneapolis, Minnesota (1977).
- Borella, H.M., J.E. Estes, C.E. Ezra, J. Scepan and L.R. Tinney, 1982, "Image Analysis for Facility Siting: A Comparison of Low- and High-Altitude Image Interpretability for Land Use/Land Cover Mapping," NUREG/CR-2961y S-744R. Prepared under Nuclear Regulatory Commission Contract No. NRC FIN A1251-2, pp.61.
- Hord, R. Michael and William Brooner, "Land Use Map Accuracy Criteria," Photogrammetric Engineering and Remote Sensing, Vol. 42, No. 5, pp.671-677 (1976).
- Rosenfeld, G.H., K. Fitzpatrick-Lins, and H.S. Ling, "Sampling For Thematic Map Accuracy Testing," Photogrammetric Engineering and Remote Sensing, Vol. 48, No. 1, pp.131-137 (1981).
- Thompson, Morris M., Ed., Manual of Photogrammetry, Vol. I, 3rd Edition, American Society of Photogrammetry, Falls Church, Virginia, pp.295-345 (1965).

POTENTIAL OF A NEW TECHNIQUE FOR  
REMOTE SENSING OF HYDROCARBON  
ACCUMULATIONS AND BLIND URANIUM  
DEPOSITS: BURIED LiF  
THERMOLUMINESCENCE DOSIMETERS

F.R. SIEGEL, Department of Geology, The George Washington  
University, Washington, D.C. 20052 USA  
J.E. VAZ, Instituto Venezolano de Investigaciones Cientificas,  
Apartado 1827, Caracas, Venezuela  
R.C. LINDHOLM, Department of Geology, The George Washington  
University, Washington, D.C. 20052 USA

ABSTRACT

Buried thermoluminescence dosimeters may be useful in remote sensing of petroleum and natural gas accumulations and blind uranium deposits. They act as integrating detectors that smooth out the effects of environmental variations that affect other measuring systems and result in irregularities and poor repeatability in measurements made during gas and radiometric surveys.

INTRODUCTION

The radiometric survey is the primary prospecting method for uranium deposits. The radiometric survey is an adjunct technique in prospecting for petroleum and natural gas accumulations and would be especially valuable in lowering the front end exploration or field extension costs.

In theory, gases generated by the radioactive decay of uranium or thorium ore ascend through the overlying rock cover and/or soil system. The gases radioactive signals are detected at the surface or from fixed wing or helicopter aircraft (with a gamma-ray spectrometer or scintillometer). When a measured radiation field exceeds the "normal" field by an arbitrarily selected concentration, the explorationist has a target area for further detailed exploration.

The depositional environment of accumulation for petroleum and natural gas precursors is one that allows for the accumulation of uranium compounds. The uranium materials associate with the petroleum and natural gas containing rocks. These hydrocarbon accumulations contain the gases from the radioactive decay process. In theory, these gases should rise through the overlying rock and soil system and give radioactive "hot spots" in a halo or fan-shape. Since the radioactivity highs are believed to mark the perimeter of a hydrocarbon accumulation, the associated low concentrations would mark the exploration targets. Figure 1 illustrates this situation.

In the case of radioactivity patterns associated with uranium and thorium accumulations, the target will manifest itself as an anomalous concentration directly above the radioactive mineralization unless structural geology characteristics (dip or fracture, a

joint-fault-crack-crevasse system), hydrodynamic conditions, or coning effects from diffusion result in "offset" anomalies (Figure 2).

The radioactivity concentration patterns that are related to petroleum and natural gas accumulations are peripheral to the accumulations or form halo-shaped targets enclosing or nearly enclosing them. The reason for the peripheral or halo-shaped target lies in the genesis of petroleum and natural gas deposits. In the subsurface, the hydrocarbons form from their precursors in a "source" rock unit. The petroleum and natural gas have a much lighter specific gravity than the enclosing rocks. If the rocks are porous and permeable (able to hold and transmit fluids), the hydrocarbons with their associated radioactive gases will be mobilized upward along the density gradient until they contact an impermeable rock unit or "trap". Accumulation will take place at the trap. Where accumulation stops, at the borders of the traps where there is porosity and permeability or cracks or fissures in the rock unit, the associated radioactive gases will escape and rise vertically to the surface where peripheral or halo-shaped radioactive concentrations in excess of the natural radioactive field will develop (Figure 3).

Radiometric prospecting for uranium and thorium ores and for petroleum and natural gas has a good theoretical base. The technique has been imminently successful in the exploration for radioactive minerals. Most major deposits have been targeted using this method especially when they are not covered by great thicknesses of rock or other overburden. In petroleum and natural gas exploration, the technique has not been as successful, and is considered unconventional. It is used only as an adjunct to the accepted approaches to exploration for hydrocarbons. The major problem that exists in radiometric prospecting is one of repeatability. If the results are not reproducible in assessing the radioactivity signals in a survey, something is wrong either in the theory or in the measuring technique or equipment. Since the theory is good, the lack of repeatability in measurement must lie in the measuring technique, timing, or the environment.

## PROBLEMS

Although excellent instrumentation has been developed to measure the transient radioactivity field (portable, vehicle mounted and airborne gamma-ray spectrometers, scintillation counters, ionization chambers, and others), geochemical prospecting based on analysis of gases (e.g., Rn) or radiometric signal from the decay products from the uranium and thorium chains may be difficult to carry out. The problem of repeatability of measurement mentioned above affects confidence in results and subsequent exploration interpretation. This problem exists if insufficient amounts of gases are collected in short term sampling, if radioactivity signals are low, or if interferences result from short term environmental variations caused by meteorological and/or seasonal processes. Such variations can occur in air temperature, soil temperature, barometric pressure, time of sampling, wind, precipitation (rain, light snow, heavy snow), position of the water table, relative humidity, soil moisture, the frozen or thawed state of the soil, diurnal cycles, solid earth tide, orientation of slope sites, and thickness of

overburden (McCarthy, 1972; Klusman and Webster (1981)).

Klusman and Webster (1981) have demonstrated some of these effects beautifully in their study of free mercury vapor emission and radon emanation in a shallow instrument vault set and sealed in weathered regolith at a single non-mineralized site. The site is about 30 miles west of Denver, Colorado, at an elevation of 2640 m. Bedrock at the site is Precambrian gneiss. For example, Figure 4 compiled from Klusman and Webster (1981) shows the seasonal trends for Hg emission at the vault site and that Hg emission may vary by an order of magnitude (from  $<0.4$  ng to  $>4.0$  ng). Figure 4 also illustrates some multivariate interactions and phase lag effects in the system. The diurnal effect on Hg emission as a function of air temperature and daylight hours is shown in Figure 5 where there is a range of ng Hg from  $<0.25$  to  $>1.5$ .

Environmental variations that affect the precision of mercury emission measurements also affect gas-coupled radioactivity signal. Thus, it is obvious that geochemical exploration based on radioactivity measurements can result in false targets or can miss targets because of environmental variations or the other factors cited previously.

#### RESOLUTION OF THE PROBLEMS

The problems of insufficient signal measurable during short term sampling of surface radioactivity, and environmental interferences that affect repeatability of measurements and hence limit the usefulness of some exploration techniques must be resolved. Also, deep weathering or burial beneath a meter or more of colluvium or other transported cover masks the radioactivity from the underlying bedrock and hinders exploration. The problems may be obviated by the use of an integrating detector that can be left in the field for long periods of time. This could be used to establish background radioactivity for a near surface soil environment. With the background radioactivity level thus determined, and following arbitrary norms to fix radioactivity concentrations that encompass local and/or regional fluctuations, the "anomaly" radioactivity concentration can be set. In geochemical exploration, a measured value that exceeds the mean plus two standard deviations is often set as the threshold value above which a measurement may be considered an anomaly.

An integrating system has been used for soil gas radon determinations in uranium exploration using  $\alpha$ -track cellulose nitrate film placed in inverted plastic cups. The cups are placed open end down in a hole 0.4 to 1.0 m deep for about three weeks. The number of  $\alpha$ -track registered for a given area in a known period of time is a function of the amount of radon at the sample site. However, these cups may be sensitive to moisture condensing in them, and rainy weather and persistent seepage downward effectively prevent Rn from reaching the point of  $\alpha$ -track registration. The unit cost of a cup and "reading" the  $\alpha$ -tracks is about \$35. if a private company is used. The user must bury the cups and recover them.

We propose that thermoluminescence dosimetry is an integrating system that will work effectively in geochemical prospecting via

radioactivity measurement for uranium and thorium mineralization and for petroleum and natural gas accumulations. When semiconductors or insulating crystals interact with ionizing radiation, electrons are displaced from their ground state, or minimum energy levels, to higher, or activated, impurity or imperfection levels. The subsequent heating of these materials permits the displaced electrons to fall to a lower energy state or their ground state resulting in the emission of energy in the form of heat and light. The light is termed thermoluminescence. The amount of thermoluminescence accumulated is a function of the amount of radiation to which a crystal has been exposed.

Thermoluminescence dosimetry to establish near surface radiation levels has not been used in petroleum and natural gas prospecting. However, the potential of thermoluminescence for uranium prospecting is not new. Nielsen and Botter-Jensen (1973) used thermoluminescence to evaluate natural radiation over rock units in Greenland. Geologists on mapping projects carried  $\text{CaSO}_4:\text{Dy}$  dosimeters in their pockets in the field for three months. The radioactivity levels measured represented integrated doses for the field areas traversed. An evaluation of the dose rates measured over different rock types (crystalline rocks, sedimentary rocks, and basalts) and their average radioactive element contents indicated that the method may be useful for large scale regional prospecting. In a study in Texas, thermoluminescence of the minerals quartz and feldspar gave a ratio of lower temperature thermoluminescence to higher temperature thermoluminescence that was consistently higher in ore and the reduced rock zone of a roll-type uranium deposit than in the oxidized rock zone (Spirakis et al., 1979).

We employed buried  $\text{LiF}$  thermoluminescence dosimeters to establish background radiation levels for environmental monitoring in northern Virginia (Siegel et al., in press). The dosimeters (3.2 x 3.2 x 0.9 mm in size) were buried at 45 cm (18") depth in waterproof plastic bags and covered with soil. Of the 101 dosimeters buried, 92 were recovered after about four months. The registered radioactivity was determined by comparing the accumulated thermoluminescence against a calibration curve which was made by exposing dosimeters to known amounts of ionizing radiation (from  $\text{Co-60}$ ) and reading their thermoluminescence outputs (with an analyzer). The buried dosimeter integrated dose rates ranged from 0.06 to 1.08 mR per day (or 2.5 to 44.5  $\mu\text{R}$  per hour), an 18 fold difference. In the study, background radiation levels were also determined with a gamma-ray spectrometer but are better defined with buried thermoluminescence dosimetry. Two anomalies were found using the dosimeters which were not indicated by the gamma-ray spectrometer data.

We may speculate why anomalies indicated by thermoluminescence were not indicated by gamma-ray spectrometry or why the correlation coefficient between the two sets of data was not better than the calculated +0.64. First, the minimum dose rate that can be registered by the dosimeters (in R per month or in the cited study, 2.5  $\mu\text{R}$  per hour) is below the detection limits of most gamma-ray spectrometers (or for that matter scintillometers). Second, the field spectrometer detects only gamma rays with energies that fall within a specific count mode so that the loss of a parent isotope affects the concentration of the daughter products in the soil system. However, the thermoluminescence dosimeters register even

the less energetic gamma rays from the decay products of the uranium and thorium series. Third, the radioactivity measurements are more accurate with the buried thermoluminescence dosimeters than with the electronic field equipment. This is because the dosimeters act as integrating detectors that dilute the effects of short-term environmental changes cited previously. Measurements made with the gamma-ray spectrometer are responsive to these changes.

There is no doubt that for the reasons cited above and another we will give that this use of buried thermoluminescence dosimeters in geochemical prospecting for uranium mineralization and hydrocarbon accumulations should be field tested. This would be especially important where gamma-ray spectrometer or scintillometer methods have been used but have failed to define a uranium deposit or an oil field. For example, at the Number Three Orebody, Ranger One, Australia, various remote sensing techniques were tested in the exploration for this  $U_3O_8$  deposit (Table 1). Some were successful, and some not (Sherrington et al., 1982). Airborne gamma radiometry and ground radiometrics gave positive responses to the Ranger One Complex. However, in the same Alligator River Province, the Jabiluka II body, the largest known uranium deposit in the world is essentially blind to all methods not involving drilling. It is also undefined on two sides (Nash et al., 1981). The Jabiluka II body would be an ideal place to test the buried thermoluminescence dosimeter technique for uranium prospecting.

The testing of the buried thermoluminescence dosimeter technique for oil and/or gas exploration could be done at any known field which has had or presently has production. Weart and Heimberg (1981) report excellent results of radiometric surveys in known producing fields using a vehicle mounted 5 1/2 foot long ionization chamber. Truck mounted gamma ray spectrometers with large detector crystals (to 2000 cubic inches) are also in use (Geoprofiles, 1982). These systems are useless in areas inaccessible to surface vehicles. Such areas are generally accessible to a ground field party. Where the radiometric survey is the principal method to find radioactive mineral deposits, it is used as part of an integrated program in petroleum exploration to lower front end costs. For example, it can be used to highlight limited areas in a large exploration zone for seismic study and thus avoid seismic work in the entire zone.

A fourth reason for establishing the technique is that since the thermoluminescence radiometric survey requires only the planting of unexposed dosimeters and later recovery of the exposed dosimeters, the survey can be carried out during periods of adverse field conditions. Also areas which are inaccessible to vehicle mounted equipment are accessible to the explorationist who can carry hundreds of the thermoluminescence dosimeters into the field. For example, in tropical rain forests, dosimeters can be planted before the rainy season begins and be collected a few months later when field conditions are favorable. At another extreme, we buried thermoluminescence dosimeters in the Antarctic during the past austral summer; these will be recovered and read next December or January.

Finally, we would like to note that the technique would seem to be cost effective and once the dosimeters are recovered, the data they carry can be obtained rapidly. The dosimeters cost about \$2.00 each and are reuseable. They can be read at a rate of about 100 per

Table 1. Remote Sensing Parameters Used in Exploration of the Number Three Uranium Orebody, Ranger One, Australia (compiled from Sherrington et al., 1982). The symbols "+", "+" and "-" summarize (with qualifications) the success of each method, from useful through equivocal to non-responsive.

AIRBORNE	Landsat Photographs - no mineralization related features	
	Aerial photography - no obvious correlation of any feature with mineralization	
	Airborne gamma radiometric surveys - give a very strong response (mineralization intersects the land surface)	
	Airborne magnetic surveys - do not delineate the bodies themselves	
GROUND	Regional gravity surveys: assist defining boundaries	
	Ground magnetics	-
	Ground radiometrics	+
	Gravity	+
	V.L.F. electromagnetics	-
	Transient electromagnetics	-
	Self potential	-
	Resistivity	-
	Surface pea-gravel geochemistry	+
	Soil geochemistry at 0.5-1 m deep	+
	1-2 m deep	+
	Soil geochemistry profiles 0-5 m deep	+
	Radon gas	+
	Helium gas	+
	Biogeochemistry	+
	Hydrogeochemistry	+
	Stream sediment sampling	+

day. The major cost in an exploration project is in manpower to plant and recover them. The cost is warranted if front-end exploration expenses are lowered, for example, in hydrocarbon exploration, or if uranium deposits are detected that can not be found by other methods.

## REFERENCES

1. Foote, R.S., 1969, Review of radiometric techniques in petroleum exploration: In W.B. Heroy, ed., *Unconventional methods in exploration for petroleum and natural gas*, Southern Methodist Univ. Press, Dallas, 43-55.
2. Geoprofiles, 1982, Radiometric oil exploration: looking for halos: EG & G Geometrics, 5, 5, Sunnyvale, Calif.
3. Klusman, R.W. and Webster, J.D., 1981, Meteorological noise in crustal gas emission and relevance to geochemical exploration: *J. Geochem. Explor.*, 15, 63-76.
4. McCarthy, J.H. Jr., 1972, Mercury Vapor and Other Volatile Components in the Air as Guides to Ore Deposits: *J. Geochem. Explor.*, 1, 143-162.
5. Nash, J.T., Granger, H.C. and Adams, S.S., 1981, Geology and concepts of genesis of important types of uranium deposits: *Economic Geology, Seventy-Fifth Anniversary Volume*, 63-116.
6. Nielsen, B.L. and Botter-Jensen, L., 1973, Natural background radiation levels from areas of major geological units in Greenland, determined by means of thermoluminescence dosimetry: *Modern Geology*, 4, 119-129.
7. Sherrington, G.H., 1977, Some aspects of natural gamma radiation in ore search: In C.R.M. Butt and I.G.P. Wilding, eds., *Geochemical Exploration 1976*, *J. Geochem. Explor.*, 8, 325-335.
8. Sherrington, G.H., Browne, A.L.L., Duffin, R.H. and Danielson, M.J., 1982, Number Three Orebody, Ranger One, Australia: A case history (abs.): In *Program with Abstracts, 9th Int'l Geochemical Exploration Symposium*, Saskatoon, Canada, 1-2.
9. Siegel, F.R., Lindholm, R.C. and Vaz, J.E., in press, In-ground Environmental Radioactivity Monitoring in the Culpeper Basin, Virginia, Using LiF Thermoluminescence Dosimeters: *Environmental Geology*.
10. Spirakis, C.S., Goldhaber, M.B. and Reynolds, R.L., 1979, Thermoluminescence as a potential uranium prospecting tool: In Watterson, J.R. and Theobald, P.K., eds., *Geochemical Exploration*, 1978, 77-80.
11. Weart, R.C. and Heimberg, G., 1981, Exploration radiometrics: post-survey drilling results: In Gottlieb, B.M., ed., *Unconventional Methods in Exploration for Petroleum and Natural Gas II*, SMU Press, Dallas, 116-123.



## Figure Captions

- Figure 1. Hypothetical section across an oil-bearing structure. The pattern of molecular movement of hydrocarbon gases from the oil mass towards the surface and the movement of subsurface water with its natural radioactive component are portrayed. There are concentrations of radioactivity and hydrocarbon gases at the surface position vertical projection peripheral to the subsurface petroleum accumulation (From Merritt, 1959).
- Figure 2. Conceptual diagram showing the distribution of uranium and  $^{214}\text{Bi}$  (eU or equivalent uranium) downslope from an oxidizing uranium deposit. Note the displaced eU anomaly where groundwater bearing radioactivity issues as a spring, and the false anomaly for uranium in the bog where sufficiently reducing conditions have resulted in the precipitation of uranium (After Bradshaw and Lett, 1980).
- Figure 3. Results of a radiometric survey at the Coyote Creek and Madison Pole Hills Field, Bowman County, North Dakota. Oil is found at about 9800 feet in the Ordovician Red River dolomite. Note the halo-shaped pattern for the radiometric high region (modified from Weart and Heimberg, 1981).
- Figure 4. Variations in Hg emission in relation to changes in soil temperature, soil moisture, and water table level as measured over a one year period. Gas-coupled radioactivity signals may be subject to the same variations (compiled from Klusman and Webster, 1981).
- Figure 5. Diurnal variation in Hg emission in relation to changes in outside temperature. Gas-coupled radioactivity signals may be subject to the same variation (after Klusman and Webster, 1981).

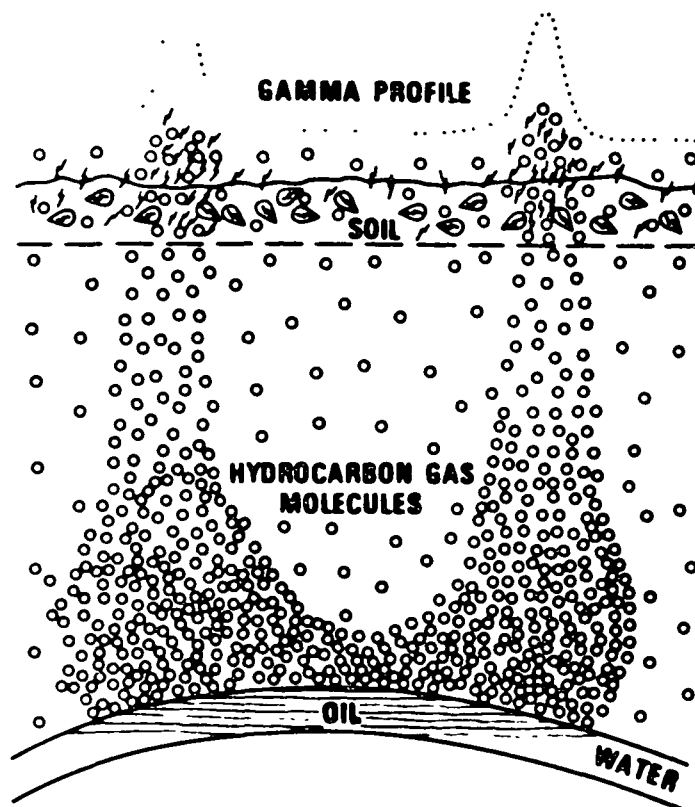


Figure 1

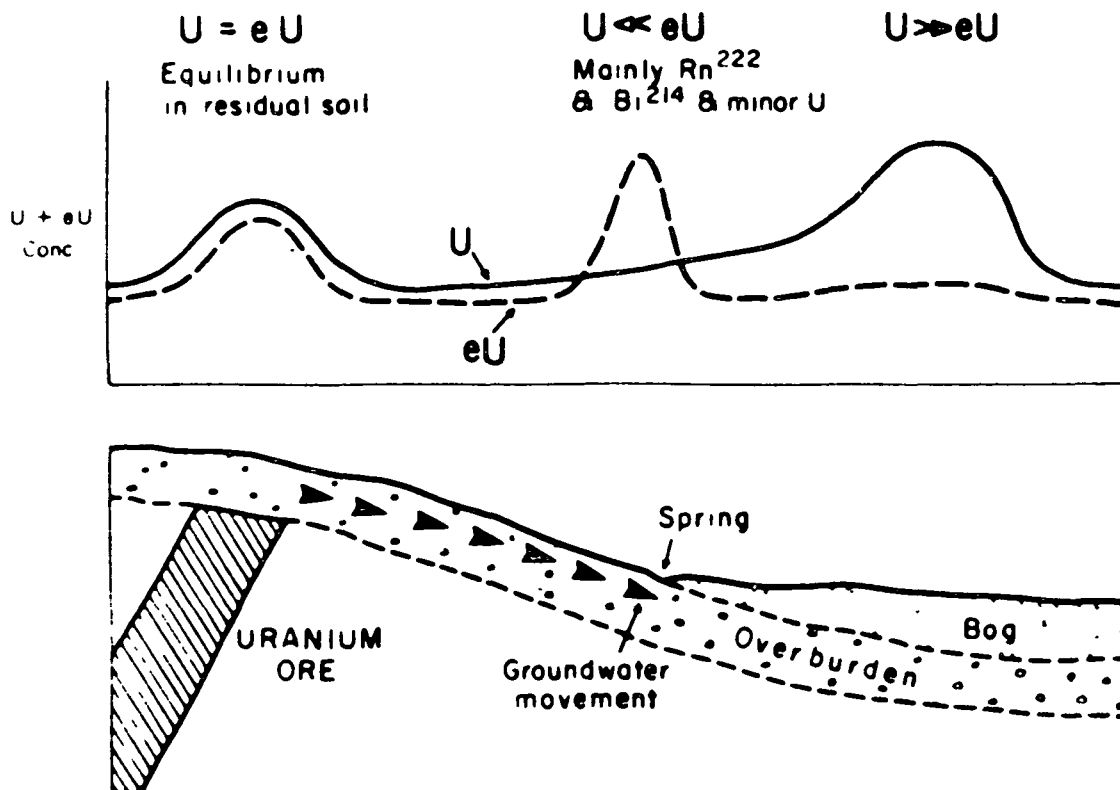


Figure 2

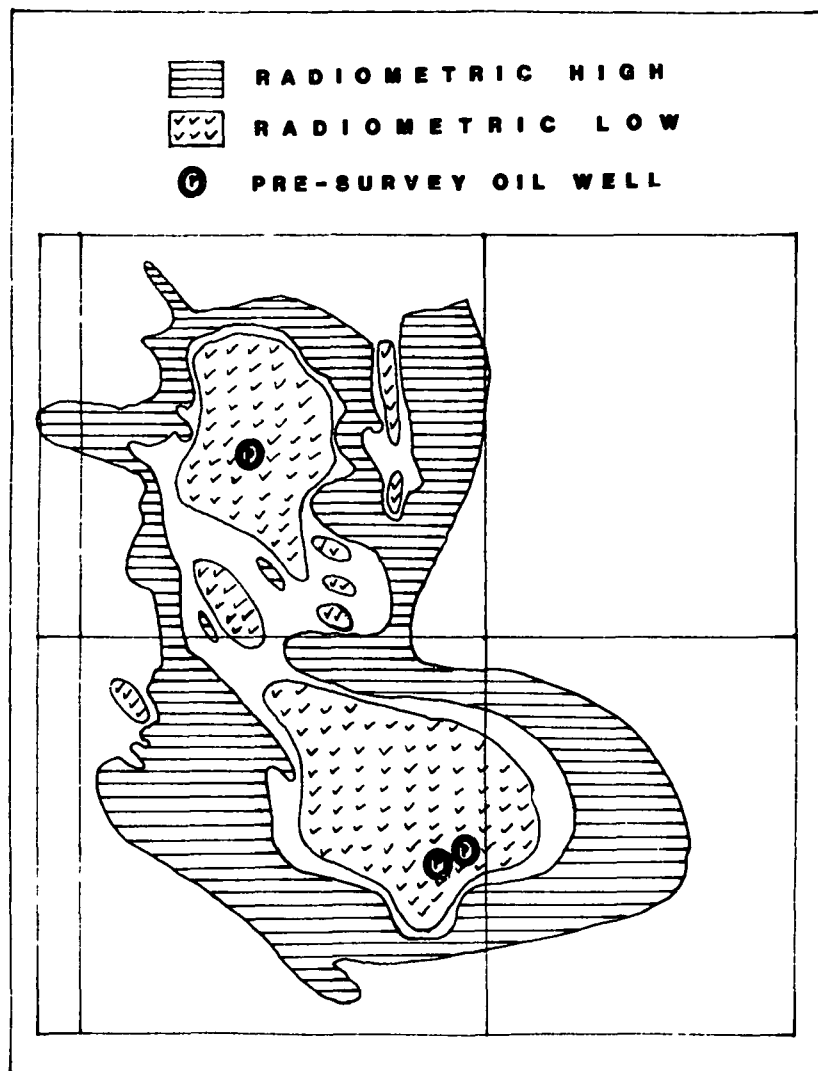


Figure 3

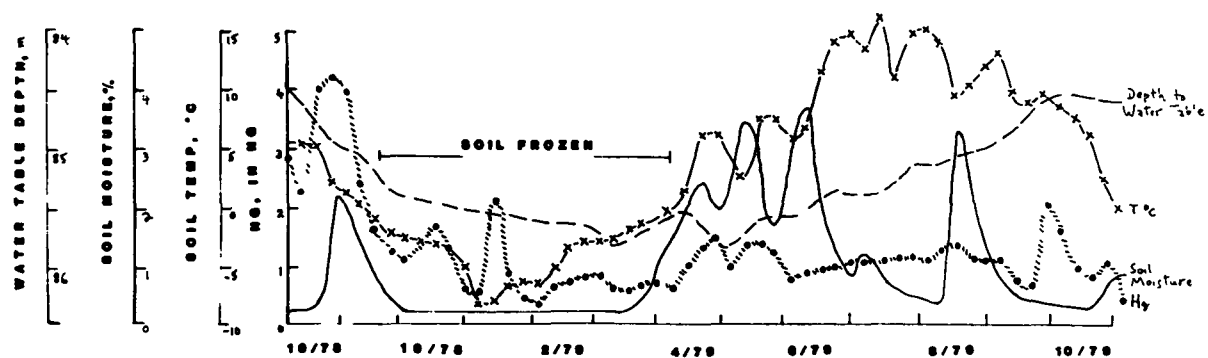


Figure 4

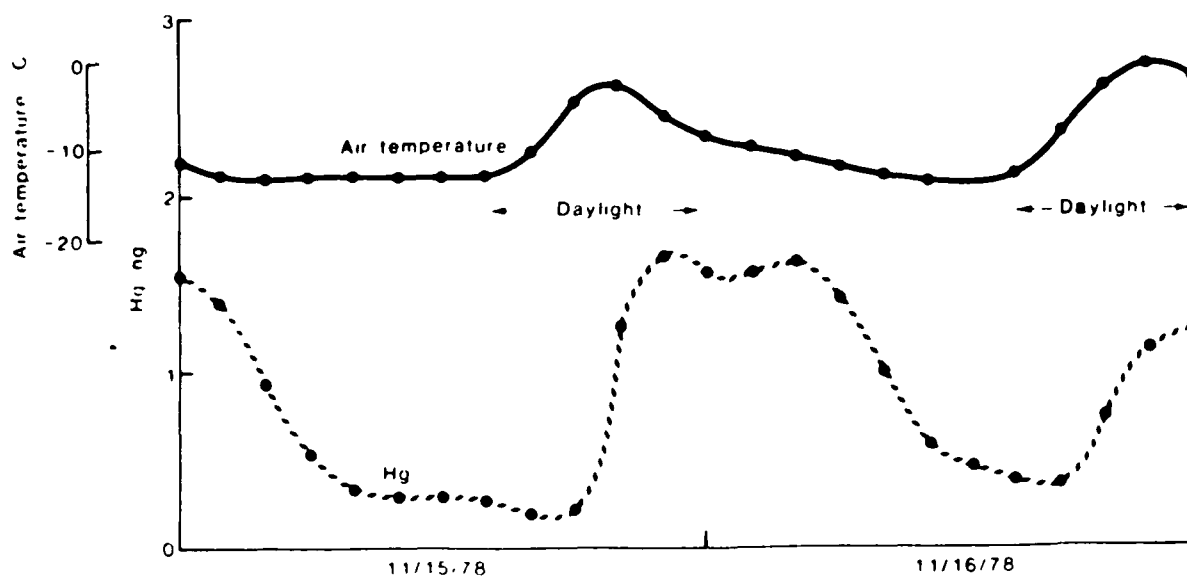


Figure 5

Part 2

INTEGRATION OF REMOTELY SENSED  
AND OTHER GEO-REFERENCED DATA BASES  
INTO GIS FOR MODELING AND APPLICATIONS

**Page Intentionally Left Blank**

GIS INTEGRATION FOR QUANTITATIVELY DETERMINING  
THE CAPABILITIES OF FIVE REMOTE SENSORS  
FOR RESOURCE EXPLORATION

R. Pascucci and A. Smith  
Autometric, Inc.  
Falls Church, VA 22047 U.S.A.

EXTENDED ABSTRACT:

To assist the U.S. Geological Survey in carrying out a Congressional mandate to investigate the use of side-looking airborne radar (SLAR) for resources exploration, a research program was conducted to define the contribution of SLAR imagery to structural geologic mapping and to compare this with contributions from other remote sensing systems. Imagery from two SLAR systems and from three other remote sensing systems was interpreted, and the resulting information was digitized, quantified and intercompared using a computer-assisted geographic information system (GIS). The study area covers approximately 10,000 square miles within the Naval Petroleum Reserve, Alaska, and is situated between the foothills of the Brooks Range and the North Slope.

The principal objectives of the research project were: 1) to establish quantitatively, the total information contribution of each of the five remote sensing systems to the mapping of structural geology; 2) to determine the amount of information detected in common when the sensors are used in combination; and 3) to determine the amount of unique, incremental information detected by each sensor when used in combination with others. The remote sensor imagery that was investigated included real-aperture and synthetic-aperture radar imagery, standard and digitally enhanced Landsat MSS imagery, and aerial photos.

Imagery from each of the five sensor systems was interpreted for evidence of geological structural features, which, within the confines of the study area consisted of anticlinal axes, synclinal axes, and lineaments that were interpreted to be the surface expression of underlying faults and fractures. Next, the overlays containing the interpretation results were digitized for entry into an automated geographic information system designed for the storage, retrieval, manipulation, and display of geographic-based information. Finally, manipulations were performed on the digital maps in the GIS data base to produce single- and multiple-theme structure maps, to compute statistical data enumerating the total number and length of structural features on the overlay from each sensor system, to measure the length of structures detected in common by two or more sensors, and to measure the length of structures detected uniquely by each sensor.

In respect to the total information content of each sensor, the principal results of the GIS manipulations were as follows: 1) the enhanced Landsat MSS detected 5876 km of structural information; 2) aerial photos

detected 5650 km (extrapolated from a smaller sample); 3) real-aperture SLAR detected 5589 km; 4) synthetic-aperture SLAR detected 3991 km; and 5) standard Landsat MSS detected 3697 km. In respect to information detected in common by sets of sensors, the results of the digital overlay and mensuration operations of the GIS showed that only about one-third was detected in common, and, conversely, about two-thirds of the structural geologic information was detected uniquely by one and/or the other sensor. The meaning of these results is that, in mapping geologic structure either for energy exploration or for power plant siting, it is far more important than has previously been thought to use two or more remote sensing systems and thereby to take advantage of the large amount of information uniquely detected by each.

The results of the remote sensor image interpretation were synthesized and used in the production of a map showing favorable hydrocarbon exploration targets.

## 1. INTRODUCTION

The House/Senate Conference Report on HR 4930 (96th Congress), Department of the Interior and Related Agencies Appropriations, 1980, states that the U.S. Geological Survey should "begin the use of side-looking airborne radar imagery for . . . geological mapping and geological resource surveys in promising areas [of the United States], particularly Alaska."

To aid the Geological Survey in this effort, Autometric, Inc. has conducted a research program to evaluate and compare the geologic information content of real- and synthetic-aperture SLAR systems and to define the contribution of SLAR and other remote sensor imagery to structural geologic mapping. In the course of this research project, imagery from five different remote sensors was interpreted, and the resulting information was quantified and intercompared using a computer-assisted geographic information system developed for the U.S. Fish and Wildlife Service.\* The imagery that was examined consisted of: real-aperture SLAR (APS/94D) imagery, synthetic-aperture SLAR (GEMS-1001) imagery, standard Landsat multispectral scanner (MSS) imagery, digitally enhanced Landsat MSS imagery and color aerial photographs.

The study area included two U.S. Geological Survey quadrangles of the 1:250,000-scale, map series: viz., the Utukok River and Lookout Ridge quadrangles in the North Slope and northern foothills of the Brooks Range. The study area lies entirely within the Naval Petroleum Reserve - Alaska, which the federal government has recently opened to exploration by private industry, with the first lease sale scheduled for later this

\* This geographic information system is marketed by Autometric, Inc. under the name AUTOGIS.



year. The use of a computer-assisted geographic information system to integrate and synthesize the structural analyses of multiple remote sensor data sets should contribute significantly to the planning and execution of exploration programs in this important area.

## 2. OBJECTIVES

The principal objectives of the research project were:

- o To establish, quantitatively, the total information contribution of each of the five remote sensing systems to the detection of structural geological features. The term "total information contribution" is defined here as the total length of structural features detected on the imagery. The sensor systems investigated were real-aperture and synthetic-aperture SLAR, standard and digitally enhanced imagery from the Landsat Multispectral Scanner, and aerial photos.
- o To determine the amount of structural information detected in common by two or more sensors in combinations of imagery from the five sensor systems. For example: when SLAR and MSS imagery of the same area is interpreted, how much of the resulting geological information is detected by both sensors.
- o To determine the amount of unique, incremental structural information detected by each sensor in combination with others. The term "unique, incremental information" is defined as the total amount of information detected by a sensor minus the amount detected by that sensor in common with other sensors. For example: when SLAR and MSS imagery of the same area are interpreted, how much of the resulting geological information is detected by each that was not detected by the other.

## 3. SCOPE AND METHODOLOGY

### 3.1 Description of the Study Area

The Utukok River/Lookout Ridge area lies just north of the Brooks Range between 69 and 70 degrees north latitude and 156 and 162 degrees west longitude (Figure 1). It contains approximately 26.156 square kilometers, virtually all of which lie within the Naval Petroleum Reserve, Alaska. The topography varies from the flat, low-lying land of the Alaska North Slope, in the northern half of the area, to the more elevated and rolling ridge-and-valley topography that has been developed on the folded strata in the southern half. The principal underlying rocks consist of marine and continental consolidated sediments of lower and upper Cretaceous age (Beikman, 1980). Except for willows immediately adjacent to streams, the vegetation consists of tundra grasses, mosses, and bushes (Chapman and Sable, 1960).

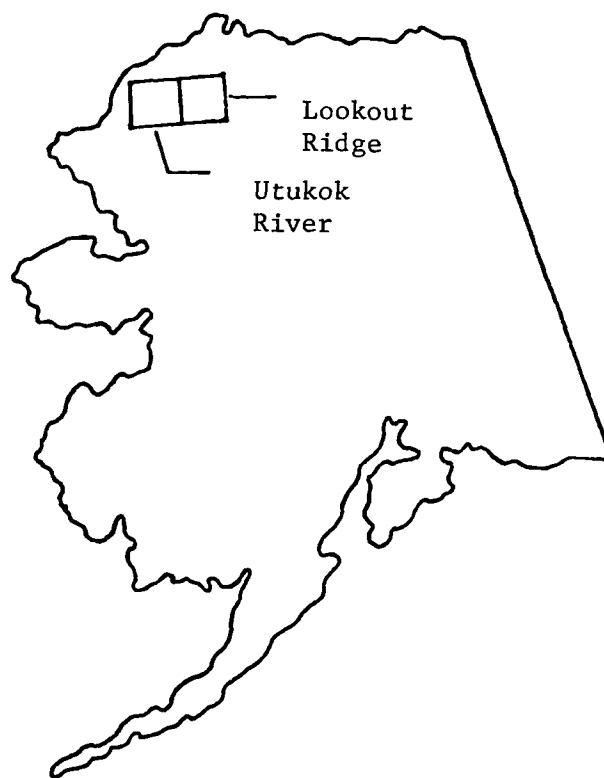


Figure 1. Location Map Showing Remote Sensor Research Area

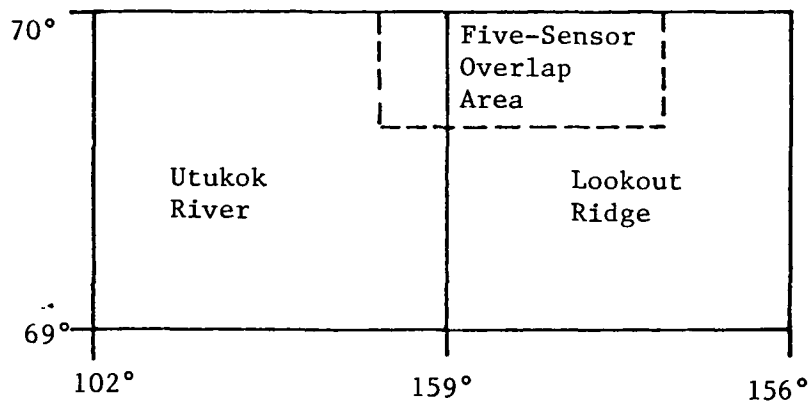


Figure 2. Location Map for Five-Sensor Overlap Area

The Utukok River/Lookout Ridge area was completely covered by the four types of SLAR and MSS imagery. Within this area, a sub-area of 8365 square kilometers was delineated for separate study. This sub-area -- called the Five-Sensor Overlap Area (Figure 2) -- is the location for which aerial photographs were available in addition to the two SLAR systems and the two kinds of Landsat imagery. The most significant aspects of the study area in respect to this investigation are its vegetation (grass) and its topography (flat to rolling). In areas characterized by arboreal vegetation or by mountainous topography, the results of this kind of study would probably be very different.

### 3.2 Interpretation of the Remote Sensor Imagery

The geologic interpretation was limited to structural features, which, within the confines of the study area, consisted of anticlinal axes, synclinal axes, and lineaments that were considered to be the surface expression of underlying faults or fractures. The interpretation of such features is relatively straightforward as compared with features such as, for example, lithologic boundaries. In most cases, a lineament or fold axis that has been detected by one geologist will be seen by another, especially if it is pointed out to him, whereas the detection of lithologic boundaries is more difficult, more subjective, and more liable to conflicting interpretations. Therefore, since the principal purpose of the research project was to measure the amount of geological information detected on remote sensor imagery, it was decided to make the interpretation of this imagery as objective as possible by restricting it to the mapping of structural features, that is, lineaments and fold axes.

Lineaments were subdivided into two categories: "possible faults" and "probable faults". In general, "possible faults" are those lineaments characterized by alignment of geomorphologic features, hydrologic features, lithologic units, vegetation, or tone. "Probable faults" are lineaments characterized by a lateral offset of the same five types of features.

Lineaments having a trend subparallel to bedding were assumed to be the surface expression of bedding planes and were not annotated as lineaments unless other evidence was present, as when two adjacent synclines were observed without an intervening anticline.

The order in which the imagery was interpreted was: SLAR first, followed by standard Landsat MSS imagery, aerial photos, and digitally enhanced Landsat MSS imagery. This may have introduced a cumulative bias in favor of each subsequently-interpreted set of imagery, since it is probable that a cumulative geological learning process took place during the course of interpretation.

It should be noted that the SLAR data acquisition program, which produced the SLAR imagery used in this research project, was not designed as a controlled scientific experiment but as a practical test of the data products of the two SLAR systems. The SLAR data acquisition contractors

were encouraged by the Geological Survey to select the mission design criteria that would, in the light of their experience, produce the best results. Thus, such design parameters as date of acquisition, flying altitude, look direction, and depression angle were not the same for both of the SLAR systems.

SLAR imagery interpretation. Both the synthetic-aperture and real-aperture SLAR imagery were subjected to separate interpretations by two remote sensing geologists working independently of one another. The system that was followed was that geologist A interpreted the real-aperture imagery first followed by the synthetic-aperture imagery, whereas geologist B began with the imagery from the synthetic-aperture system and went on to that from the real-aperture system. When both geologists had completed their interpretations, they then placed both of their transparent overlays on the imagery and discussed each delineation that was not common to both. In by far the greater number of cases, the geologist who had not delineated the feature had simply overlooked it, but in some instances there was a good deal of discussion as to whether the feature was distinct enough or linear enough or long enough to qualify as a fault or fracture. Following these discussions, a composite overlay was synthesized which contained all the interpretations upon which both investigators had agreed. This same system was adhered to in the interpretation of all five imagery data sets.

Landsat MSS imagery interpretation. Two different types of Landsat imagery were interpreted in this project: (1) standard (unenhanced) off-the-shelf Landsat MSS products at a scale of 1:500,000, and (2) digitally enhanced (contrast enhanced and edge enhanced) Landsat MSS products at a scale of 1:250,000. Both types of Landsat products were prepared by the EROS Data Center.

Standard Landsat MSS imagery covering the Utukok River/Lookout Ridge area was interpreted. For both areas, the investigators utilized coverage that consisted of black-and-white (bands 5 and 7) and color IR imagery, all at a scale of 1:500,000. Since complete overlapping coverage was available, the interpretation was performed both stereoscopically and monoscopically. Imagery from two seasons (April and July) was interpreted in order to take advantage of the additional information that might result from different azimuths and elevations of solar illumination and from different surface coverings (snow in the April scenes and tundra vegetation of grass and moss in July). Although no measurements were made, it was apparent that much more information was derived from the April scenes, probably due to the lower sun elevation illuminating an unbroken and spectrally uniform cover of snow. The deep, uniform red reflectance of the July vegetation tended to mask the tonal variations caused by topography.

Digitally enhanced Landsat MSS products were also interpreted by the investigators. These products were prepared at the EROS Data Center using their in-house digital image processing systems and EROS Digital Image Processing (EDIPS) tapes covering the study area.

One of Autometric's remote sensing geologists traveled to the EROS Data Center to work with EROS staff members in preparing the digitally enhanced products. The most important part of the effort consisted of using the interactive digital image analysis equipment and techniques available at the EROS Data Center to prepare, review, evaluate and select the optimum contrast stretches and edge enhancements for each of the four Landsat tapes.

Upon completion of the interactive image analysis sessions, the four tapes were processed for preparation of 1:250,000-scale, black-and-white prints in bands 5 and 7. These prints were then interpreted using the same geologists and methodology as were employed in interpreting the SLAR and standard Landsat MSS imagery.

Interpretation of aerial photography. Since it was not possible, within the constraints of time and budget, to interpret large-scale aerial photographic coverage of the entire two-quadrangle area, stereoscopic aerial photos were interpreted at approximately the same level of effort as was used in the interpretation of the SLAR and Landsat imagery. This involved the stereoscopic interpretation of 87 color photos at an average scale of approximately 1:80,000. The photos (taken in June, 1971) cover the Five-Sensor Overlap Area located in the north-central portion of the Utukok/Lookout area. (See Figure 2.) Linear remnants of snow in topographic depressions were a considerable aid to the interpretations.

Interpretation of Seasat SAR. Seasat radar imagery was interpreted within the Five-Sensor Overlap Area, at a scale of 1:500,000. The photographic quality of the imagery was very poor, however, and it was felt that the information derived from it by interpretation was far less in quality than that which would have been derived from a more typical image sample and was certainly far less than that which would have been derived from a digitally processed scene. Thus, since previous experience of the investigators indicated that this particular Seasat radar imagery was not at all representative of its true performance capability, the results of the Seasat imagery interpretation have not been included in this report. This is especially regrettable, since Ford (1980) detected twice as many lineaments on Seasat SAR than on standard Landsat MSS in a study area in the Appalachians, and corroboration in the treeless environment of northern Alaska would have been most interesting.

### 3.3 Digitization and Manipulation of the Interpreted Data

Upon completion of the interpretation, the overlays were digitized for entry into the geographic information system, a system designed in part by Autometric personnel and installed at the U.S. Fish and Wildlife Service facilities in Fort Collins, Colorado.

This system, marketed and installed by Autometric as the Automated Geographic Information System (AUTOGIS), is a computer software system that was specifically designed for the input, storage, retrieval, manipulation,

and display of map-based geographic information. The system is scale-independent; thus, maps at any scale can be input for comparison and/or analysis. The two principal subsystems are AMS (Analytical Mapping System) and MOSS (Map Overlay and Statistical System).

The subsystem used for digitization - AMS - allows geographic information to be digitized from maps or remote sensor imagery and to be stored in a topologically valid form in a geographic data base for subsequent map generation at any scale.

The completed interpretation overlays were digitized using a standard X-Y digitizing table. Once this process had been completed, the data set was used as input into a verification process that checked the spatial consistency of the data set. Editing capabilities were then utilized as necessary to add, replace, or delete any topologically inconsistent data.

The features on the interpretation overlays are treated by AMS as a set of discrete points organized in such a way as to form line segments. Data entry and storage in AMS is organized on a "geounit" basis, which is defined simply as a "rectangular" parcel of area on the earth's surface. In this project the geounits were two 1 x 3-degree areas coincident with the standard USGS quadrangle sheets covering the project area.

Each input overlay was digitized and stored as an individual map in the geographic data base.

### 3.4 Map Production and Digital Manipulation

Once the interpretation overlays had been digitized and edited, the digital data were stored in the geographic data base in a form suitable for quick and efficient retrieval and analysis. The software system used for this purpose - MOSS - allows the user to perform a large number of functions related to map preparation, synthesis and analysis. The three principal sub-tasks of this effort were: (1) automated production of single- and multiple-theme structural maps at a common scale of 1:250,000; (2) the computation of statistical data concerning the total number and length of structural features shown on each map; and (3) measurements of the information detected in common by two or more sensors.

Map production. In the first sub-task, a variety of single and multiple data set maps were produced in order to assess the spatial relationships between structural features that had been delineated on different input interpretation overlays. (Fifty-one maps were produced as part of this project.) By using the computer-assisted mapping system, it was possible to compile maps at any desired scale showing any desired combination of sensors (radar plus Landsat, or radar plus Landsat plus photos, etc.) or geologic features (faults or folds, or faults plus folds).

Map preparation consisted of using a "CALCOMP" program designed to

allow the user to develop any desired map from data that have been digitized previously. Factors to be considered in using the program are the type of input data, the scale of the output map, the themes to be displayed, and the line symbology and color to be used for each theme.

Measurement of total length and number of structural features. The second sub-task of the digitization and manipulation task was the computation, by MOSS, of the total number and length of each kind of geologic feature shown on each interpretation overlay. MOSS capabilities include an interactive program that allows the user to display on a CRT terminal certain characteristics of each input map. Since these data were originally entered during the digitization process, their recall and display were relatively simple and rapid. Table I is a simulation of a typical display of Landsat statistical data for the Lookout Ridge map sheet. It shows that 2431.6 kilometers of geographic features were mapped, in the categories of: probable faults (69.3 km and 26 features); possible faults (1425.6 km and 233 features); synclinal axes (489.0 km and 9 features); and anticlinal axes (447.7 km and 7 features). A "PLOT" program allows the user to produce, display, and copy a small-scale CRT version of the input map. Similar statistics were acquired for each of the overlays stored in the data base.

TABLE I  
EXAMPLE OF STATISTICAL SUMMARY SHEET

Length Summary for Lookout Ridge: Standard Landsat MSS

SUBJECT	LENGTH	FREQUENCY	% TOTAL LENGTH
1. Probable Faults or Fractures	69.3 km	26	2.85
2. Possible Faults or Fractures	1425.6	233	58.63
3. Synclinal Axes	489.0	9	20.11
4. Anticlinal Axes	447.7	7	18.41
TOTAL	2431.6	275	100.00

Measurement of structural features detected in common. The third sub-task consisted of measuring the relative agreement in geologic information ("commonality") that resulted when two or more interpretation data sets from the same map sheet were combined. For example, when the "Lookout Ridge: Synthetic-Aperture SLAR" map sheet was combined with the "Lookout Ridge: Landsat" map sheet, a certain number of structures (lineaments and fold axes) were detected in common by the two sensing systems and therefore overlapped one another for some specific length. This overlap portion reflected the extent of "commonality", or agreement, between the two data sources.

By subtracting the commonality length from the total length of

geologic structures on each data set, it was possible to determine the length of uniquely detected data in each data set. That is, as defined here, "unique data" equals "total data" minus "common data".

#### 4. RESULTS

In the following sections, the information content of the sensors is discussed in terms of total contribution, common contribution, and unique contribution.

##### 4.1 Total Information Contribution of Each Sensor

Table II shows the total length of structures detected by each sensor in each of the four categories of structural geological information. The overall sensor performance, as shown by the table is: enhanced Landsat MSS detected 5876 km of structural information; real-aperture SLAR detected 5589 km; synthetic-aperture SLAR detected 3991 km; and standard Landsat MSS detected 3697 km. (If the total length of structural elements detected by aerial photos in the 8365-square-km Five Sensor Overlap Area be extrapolated over the entire 26,156-square-km Utukok/Lookout Area, it totals 5650 km, which would rank aerial photos between enhanced Landsat and real-aperture SLAR.)

Table III emphasizes the relative total information contribution of the sensors by comparing them with one another. Thus, if the enhanced Landsat is rated 100%, the contribution of the real-aperture system is 95% as much, the contribution of the synthetic-aperture is 68%, and the standard Landsat contribution is 63%. (Using the total extrapolated from the Five-Sensor Overlap Area, the contribution of aerial photos is 96% that of enhanced Landsat.)

A word of explanation should be given here concerning the great disparity in the performances of the real-aperture and synthetic-aperture SLAR systems. The real-aperture system has a resolution of 50 x 150 meters, while the resolution of the synthetic-aperture system is 10 x 12 meters, yet the real-aperture system contributed 40 percent more information. (Using the same SLAR systems in a geomorphological study conducted on the Alaska Peninsula, Cannon (1981) found that the real-aperture system contributed 25% more landform information -- 263 versus 210 landform units -- than did the synthetic-aperture radar.) The explanation for the superior performance of the lower-resolution system appears to be that, in the flat and rolling terrain of the Utukok-Lookout study area, the synthetic-aperture system with the large depression angle used (30° inboard and 11° outboard) produced a nearly shadowless image that contained less geologic information than the imagery produced by the real-aperture system with its depression angles of 21° inboard and 8° outboard. (The function of shadowing on SLAR imagery is twofold: large shadows obscure information, while smaller shadows enhance it.) It is necessary, therefore, to design the acquisition mission so that optimum shadowing is



TABLE II  
TOTAL INFORMATION CONTRIBUTION

AREA	SENSOR	Structural Element by Length (km)				TOTALS
		Probable Fault or Fracture	Possible Fault or Fracture	Synclinal Axis	Anticlinal Axis	
Utukok River/ Lookout Ridge	Synthetic Aperture SIAR	18	2660	623	690	3991
	Real Aperture SIAR	159	3063	1088	1279	5589
	Standard Landsat MSS	77	1923	814	883	3697
	Enhanced Landsat MSS	167	3984	822	903	5876
Five-Sensor Overlap Area	Synthetic Aperture SIAR	8	1239	24	63	1334
	Real Aperture SIAR	3	1238	278	298	1817
	Standard Landsat MSS	5	958	188	175	1326
	Enhanced Landsat MSS	60	1556	126	204	1946
	Aerial Photos	0	1670	76	61	1807

achieved.

Of the relative sensor performance in the Five-Sensor Overlap Area, shown at the bottom of Table III, it can be seen that enhanced Landsat again contributed the most information, based on total length, but is approximately equal to the real aperture system and aerial photos. It is also interesting to observe that, in this case, standard Landsat is approximately equal in geologic information content to the synthetic aperture system.

#### 4.2 Common Information Contribution of Each Sensor

Note that Table II addresses the total length of structural elements contributed by each individual sensor, without regard to overlaps

TABLE III  
RELATIVE INFORMATION CONTRIBUTION

AREA	SENSOR	Total Length of Structural Elements (km)	Relative Performance
Utukok River/ Lookout Ridge	Enhanced Landsat	5876	-
	Real Aperture SLAR	5589	95% of Enhanced Landsat
	Synthetic Aperture SLAR	3991	68% of Enhanced Landsat
	Standard Landsat MSS	3697	63% of Enhanced Landsat
Five-Sensor Area	Enhanced Landsat MSS	1946	-
	Real Aperture SLAR	1817	93% of Enhanced Landsat
	Aerial Photos	1807	93% of Enhanced Landsat
	Synthetic Aperture	1334	69% of Enhanced Landsat
	Standard Landsat MSS	1326	68% of Enhanced Landsat

(commonalities) in sensor contributions. The table does not show the five sensors, nor does it show the length of features uniquely detected by each sensor. This is an important consideration, for the following reason: it was seen that, in the Five-Sensor Overlap Area, the information contributed by aerial photos and real-aperture SLAR was approximately equal, for all practical purposes. Hypothetically, if the photos and the SLAR both detected the same structural geological features, there would be little reason for acquiring SLAR in areas in which aerial photos currently exist. The extent to which SLAR is required in structural geologic mapping depends on the amount of unique information it contributes.

Figures 3 and 4 show the length and percent of overlapping structural elements (features detected in common by two or more sensor systems) when data sets are compared by digital manipulation.

This set of results is very significant, showing, as it does, that when two remote sensors acquire data over a common area, only from 20 percent (aerial photos and real-aperture SLAR) to 38 percent (enhanced Landsat MSS and synthetic-aperture SLAR) is detected in common by both sensors.

A corollary to the finding that so little (about one-third) information is detected in common is the fact that a large amount (about two-thirds) of remote sensor information is detected uniquely by each sensor.

#### 4.3 Unique Information Contribution of Each Sensor

As mentioned earlier, the unique information contribution of each sensor is defined as its total information contribution minus the

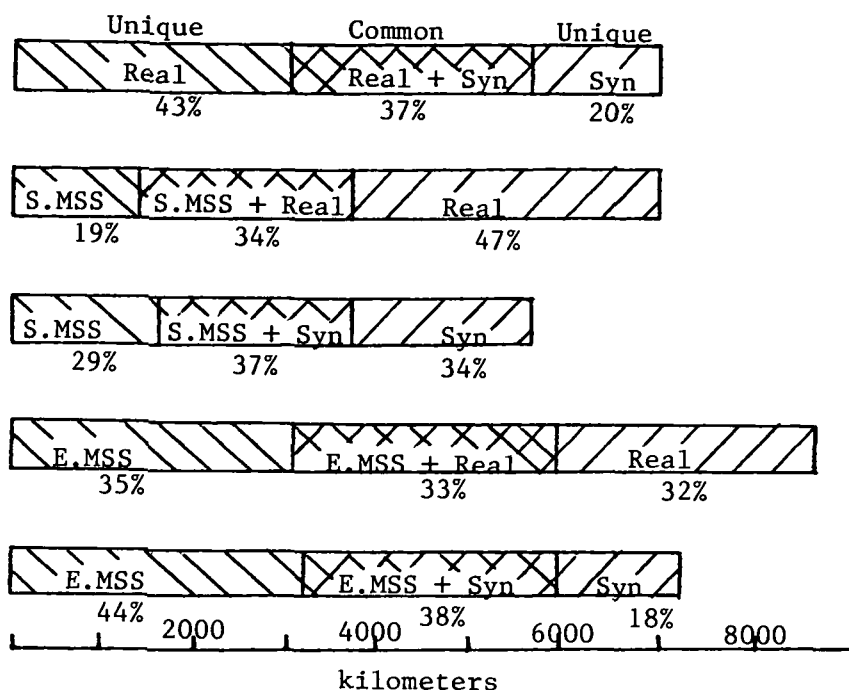


Figure 3. Percent of Common and Unique Information Contributions Utukok/Lookout Area

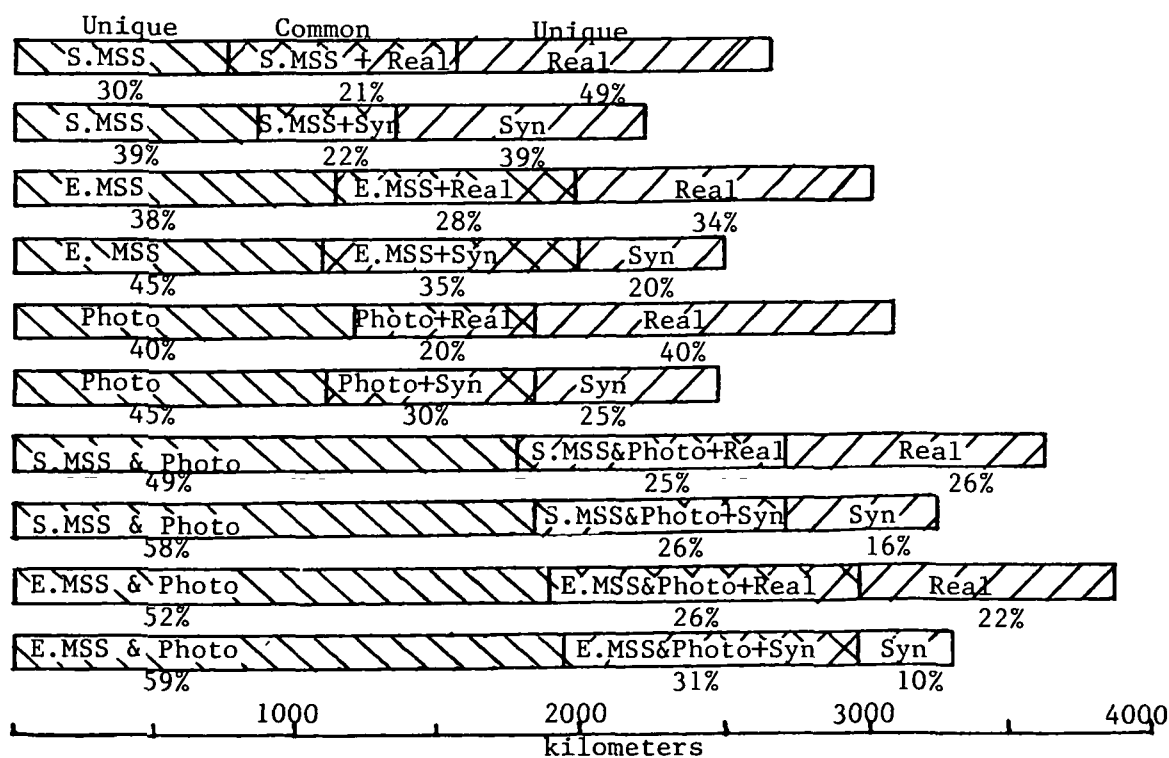


Figure 4. Percent of Common and Unique Information Contributions Five-Sensor Overlap Area

information contribution it has in common with another (or other) sensor(s). The most important finding of the investigation is the unexpectedly large amount of information contributed uniquely by each remote sensor. The application of this knowledge is obvious in such energy-related fields as resource exploration and nuclear power plant siting, where geologic structure is frequently the most important single factor to be considered. By pointing out the amount of incremental information that can be expected from the acquisition and interpretation of additional imagery, a basis is provided for more accurate cost-benefit analyses.

Figures 3 and 4 show the length and percent of the unique incremental structural information that was obtained when remote sensor data sets were used in combination.

#### REFERENCES

- Biekman, H.M., compiler. 1980. Geologic Map of Alaska, State of Alaska. Division of Geologic and Geophysical Surveys, scale 1:2,500,000, two sheets.
- Cannon, P.J. 1981. in Evaluation of Radar Imagery for Geologic and Cartographic Applications. G.K. Moore and C.A. Sheehan compilers, USGS Open-File Report 81-1358, p.20.
- Chapman, R.M., and Sable, E.G.. 1960. Geology of the Utukok-Corwin Region, Northwestern Alaska, part 3, Areal Geology. USGS Professional Paper 303-C, p.61.
- Ford, J.P. 1980. Seasat Orbital Radar Imagery for Geologic Mapping: Tennessee-Kentucky-Virginia, AAPG Bulletin, Vol. 64/12, Dec., p.2064.

THE ORSER LANDSAT DATA BASE  
OF PENNSYLVANIA

B. J. Turner

Office for Remote Sensing of Earth Resources  
The Pennsylvania State University  
University Park, Pennsylvania 16802, U.S.A.

and

D. L. Williams

NASA/Goddard Space Flight Center  
Earth Resources Branch/923  
Greenbelt, Maryland 20771, U.S.A.

1. INTRODUCTION

A mosaicked Landsat data base for Pennsylvania has recently been installed at the Computation Center of The Pennsylvania State University. Initially constructed by Penn State's Office for Remote Sensing of Earth Resources (ORSER) for the purpose of assisting in state-wide mapping of gypsy moth defoliation, the data base will be available to a variety of potential users. It will provide geometrically correct Landsat data accessible by political, jurisdictional, or arbitrary boundaries.

2. FOREST DEFOLIATION ASSESSMENT PROJECT

Each year, state and federal agencies spend millions of dollars developing programs to prevent the spread of the gypsy moth caterpillar (Lymantria dispar), which has defoliated millions of hectares of hardwood forest. Since the caterpillar was introduced in the United States in 1869, (in an effort to produce a new variety of silkworm) the gypsy moth has become established throughout most of the northeast, and south to West Virginia and Maryland. Gypsy moth populations have periodically increased to epidemic proportions. Currently one of the largest recorded outbreaks seriously infested nearly 4 million hectares (10 million acres) during the 1981 summer feeding cycle, and projections for 1982 are even higher.

Integrated pest management programs, developed to prevent the insect's spread, depend largely on accurate, timely, and efficient methods of detecting and mapping incipient forest canopy damage. Ground surveys, aerial sketchmapping, and photointerpretation have been used to detect the damage, but the expense and subjectivity of these methods have led to a search for more efficient and accurate techniques. In view of the wide areas of damage, it has also become desirable to standardize the methods used among the various state agencies.

Researchers began to look for a new survey technique which could provide timely, accurate, and standardized assessments at a reasonable cost. By the mid-1970's, after Landsat multispectral scanner (MSS) data became widely available, research began to indicate that Landsat data had potential for monitoring

widespread forest disturbances such as infestations of the gypsy moth and other insect species. The standardized spectral, spatial, and temporal coverage of Landsat data sets, and the synoptic coverage provided, seemed to be ideally suited as a survey medium. ORSER and NASA (National Aeronautics and Space Administration) at Goddard Space Flight Center (GSFC) in Greenbelt, Maryland, were among the early participants in such research.

In order to demonstrate the usefulness of satellite remotely sensed data for monitoring insect defoliation of hardwood forests in Pennsylvania, a joint research project was initiated between NASA/GSFC and the Pennsylvania Bureau of Forestry, Division of Forest Pest Management (DFPM). A framework for automated assessment of defoliation using Landsat MSS data was provided by the earlier GSFC work (Williams and Stauffer, 1978; Williams et al., 1979; Nelson, 1981).

The procedure for defoliation assessment requires four steps:

Creation of a healthy forest classification mask: Prior to insect infestation, data for a cloud-free summer Landsat image over the study site are obtained. This image is classified into two categories, using digital analysis techniques: forest and non-forest. Pixels classified as forest are assigned the value of 1 and all other pixels in the scene are assigned the value of 0. The resultant image is called the "1/0 forest/non-forest mask."

Application of the forest/non-forest mask to the image showing defoliation: An image of the study site obtained at the peak of defoliation or shortly thereafter is digitally registered to the 1/0 forest/non-forest mask. This registered image is then multiplied by the forest/non-forest mask to produce a "defoliated forest image," in which areas in the scene which show forest have been isolated from other cover types.

Application of the ratio vegetation index to assess forest disturbance: The ratio vegetation index (RVI) is the ratio of the infrared to the red spectral response (MSS band 7/band 5) for each pixel within the image. The RVI is applied to the defoliation image, creating a new image, the "assessment image," in which low ratio values indicate heavy defoliation and high values indicate healthy forest. Because of previous application of the mask, zeros indicate non-forest.

Separation of defoliation levels: Aerial surveys or other ground reference data are compared to the assessment image to determine the numerical levels separating healthy, moderately defoliated, and heavily defoliated forests. It is important to note that the key requirement in this procedure is the ability to register several different images to a common reference base. Such a common reference base has been created for the state of Pennsylvania by the Office for Remote Sensing of Earth Resources, at The Pennsylvania State University.

### 3. CREATION OF THE PENNSYLVANIA DATA BASE

The Pennsylvania legislature has mandated that the state's Division of Forest Pest Management conduct annual assessments of insect-related damage to forests throughout the state. Yearly statistics must be compiled to study trends in insect population dynamics, as well as for planning management alternatives. Although a wealth of information has been acquired over the years, it is of

limited use because it exists in various hard copy formats (e.g., maps, aerial photographs) which do not lend themselves to computer storage and retrieval, and because the non-standardized format of these products, and the subjectivity of analysis procedures used to generate them, makes meaningful trend analysis almost impossible. Landsat, on the other hand, offers a standardized MSS data source which has been collected for over 10 years. The information is in digital format, which can be processed quantitatively and repeatedly, and both the original data and the derived results can be readily stored, retrieved, and compared by computer. However, the size of the state, and the corresponding volume of data required for accurate defoliation assessment presented a unique challenge. Not only was it necessary to store and retrieve the data, but extensive digital image processing was required, as well as a means to compare and assess the output products from such processing.

In the course of the joint project between NASA/GSFC and DFPM, various methods were considered for handling the large volume of Landsat data required to conduct defoliation assessments on an annual basis. It was decided to develop a Landsat-derived, multilayered, geographic data base which could be interfaced with image analysis software. This data base had to contain a minimum of three layers:

- 1) a Landsat digital mosaic of Pennsylvania exhibiting no defoliation and registered to the Universal Transverse Mercator (UTM) map projection, rotated to north, and resampled to 57 meter square cells (the cell size of future Landsat data);
- 2) a forest resources map (forest/non-forest mask) derived from the Landsat data in the first layer; and
- 3) digitized Forest Pest Management District boundaries and county boundaries registered to the Landsat mosaic.

The capability to add additional data layers, such as the most recent Landsat data depicting defoliation, was also required.

Fortunately, the ability to retrieve, digitally process, and store Landsat MSS data sets was already available at the Office for Remote Sensing of Earth Resources (ORSER), located at The Pennsylvania State University. Thus, it was decided to develop and house the Pennsylvania Landsat data base on the IBM 370/3081 computer at the University's Computation Center. ORSER agreed to develop or acquire, upgrade, and implement all software necessary to create and manipulate the data base.

#### 4. CREATION OF THE MOSAIC

The Pennsylvania mosaic of Landsat data acquired prior to defoliation would provide the foundation for all subsequent procedures in operating a defoliation assessment system. Because of their demonstrated capabilities in generating Landsat mosaics of California and Arizona (Zobrist and Bryant, 1979), NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California was asked to generate the initial mosaic. The mosaicking procedures required the use of the VICAR/IBIS software system developed at JPL, as well as additional mosaicking software which has been incorporated into the VICAR system.

Mosaicking begins with the selection of several ground control points within each image frame. Seam control points on adjacent frames are then selected by automatic correlation analysis. These are adjusted by a distortion model for each frame, based on the ground control points. Seam points are then reconciled by averaging their mapped locations in adjacent frames. Finally, the processed Landsat data are "cut" at the mapped seam boundary to produce the mosaic piece and the pieces are "sewn" together (Zobrist et al., unpublished manuscript). The control points selected for one Landsat spectral band can be applied to the other three bands and the same geometric correction performed.

## 5. REFORMATTING THE DATA BASE

As supplied to ORSER, the magnetic tapes containing the mosaic were in band-sequential VICAR format. That is, each file contained data for a quadrangle of one degree of latitude by two degrees of longitude. Eight such quadrangles were necessary to cover the whole state. Unfortunately, this format was not suitable for the Penn State computing environment, where it is much less expensive to locate the beginning of a file on a tape than it is to read individual records. Thus, it was more efficient to store the data in long-line records, with relatively few records per file, than in large quadrangles of data. It was also more convenient to store the data in a form similar to the ORSER raw data (RD) format, a modified band-interleaved-by-line format, than in the band-sequential format.

The ORSER data base (DB) format, like the RD format, is also a band-interleaved-by-line format. Here all the pixels for one band of a scan line are stored as one logical record and the scan lines are organized in ascending order, just as in the RD format. Scan lines are grouped into files containing 500 lines. Thus, 12 files, containing 500 lines each, are used for each half of the data set. Header information on the files is stored within the program so that only the files containing data within the area of interest need be read. This reduces the computer time required to access an area that may be several thousand scan lines down the data base.

Three programs were needed to reformat the half-state data from the 16 VICAR files into the ORSER DB format: SEW reads up to four VICAR-format files of adjacent areas and concatenates them to form one VICAR file. This is done for each of the four bands. INT reads VICAR files and generates band-interleaved-by-line files. It is run on bands 4 and 5 together and then on bands 6 and 7 together. DBGN then reads these two files, interleaves them, and breaks them down into 12 files of 500 scan lines each. To check the results, band 7 of the complete data set was displayed on a Versatec electrostatic printer (Figs. 1 and 2). The three reformatting programs can also be used to add information to the data base, such as extra bands of Landsat data or data for adjacent geographic areas.

In addition to the grid-cell formatted Landsat data, the data base consists of sets of coordinates, stored on separate tape files, describing irregular areas, such as the county and forest district boundaries currently in the system. An index in the front-end system relates each county and forest district name to its corresponding file on the tape. Additional boundaries (watersheds, for instance) can be added to the system, as long as their coordinates are in the UTM projection.



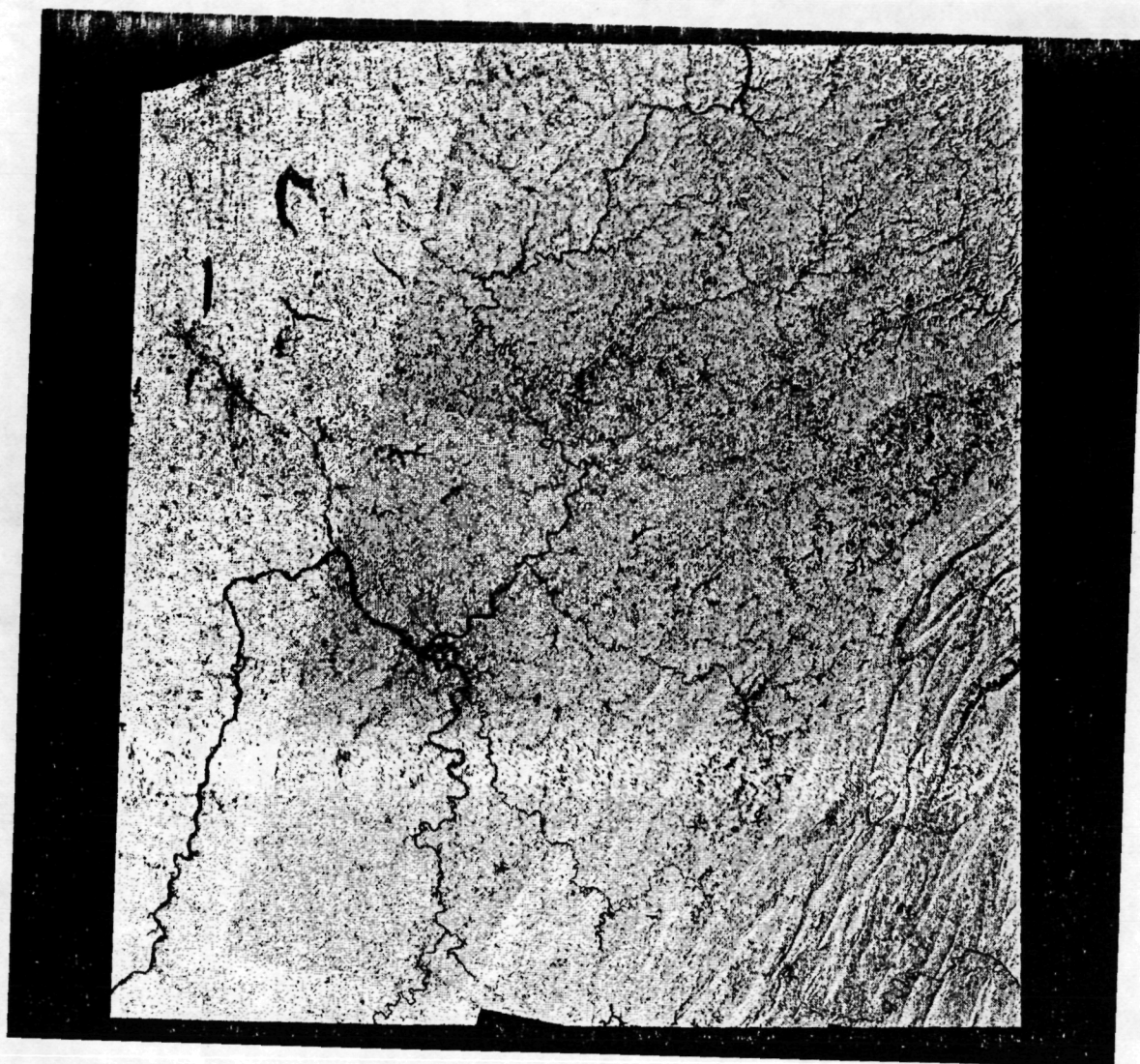


Figure 1. Band 7 electrostatic printer display of the western half of the Pennsylvania mosaic (UTM 17).



Figure 2. Band 7 electrostatic printer display of the eastern half of the Pennsylvania mosaic (UTM 18).

## 6. SUBSETTING FROM THE DATA BASE

The SUBDB program was written to subset an irregularly bounded area from a data set in the ORSER DB format and output it in the RD format for subsequent analysis by any of the ORSER programs. The program may read a file containing the coordinates describing a predefined polygon, or coordinates may be entered directly in UTM meters or as line-and-element numbers. These coordinates are converted to start and stop points within each scanline. The program then determines which file to start with, processing sequentially from that file. The data are reformatted and all pixels lying outside the defined polygon are replaced with zeros (null pixels). The new data set, now in the ORSER RD format, is written to tape and can be processed by any ORSER program that handles raw data. An example of a county data set extracted in this manner from the data base and displayed on an electrostatic plotter using the NMAP program is shown in Fig. 3. In order to extract the UTM coordinates of counties and forest districts supplied on tape from GSFC, the PIOS program was written. It converts UTM coordinates to line-and-element numbers, producing input to the SUBDB program.

## 7. DEVELOPMENT OF THE FRONT-END

In most cases, using the data base involves moving large data sets between storage media and the computer. Because such transfers require manipulating job control language (JCL)--a process unfamiliar to many potential users--a user-friendly front-end processor was developed to set up jobs. At the University, where the large IBM 370/3081 operates in batch mode, the best way to develop such a front-end was to use the EXECUTE facility of the INTERACT (also known as MENTEXT or WYLBUR) system introduced at the University two years ago.

The INTERACT system is designed for program development, remote job processing, and document composition. Responding to commands from local or remote terminals, it interprets these, performs the requested processing, prompts for further information, and provides error messages where appropriate (Cullinane Corp., 1980). Using the EXECUTE facility of INTERACT, which provides a complete programming language, the user can construct an EXECUTE (EXEC) file, containing an executable series of instructions. Such files are commonly used by non-programming personnel to perform operating system functions, and are particularly useful for handling frequently-used functions involving data manipulation. The INTERACT front-end for the ORSER system has proven very useful for users at the University (Turner et al., 1982).

To operate the Pennsylvania data base, an EXEC file was set up as a major subset of the existing ORSER EXEC file. After entering the ORSER EXEC file, the data base user responds to the first prompt by typing in "DATABASE." A series of prompts then permits the user to select the county, forest district, grid cell, quadrangle area, or irregular polygon desired; asks for the name, number, or coordinates of the specified area; and asks for the band numbers required, and whether the output is to be put on tape or disc. By typing in "HELP" to any of these prompts, the user is supplied with further explanation of the reply appropriate to that prompt. The result of the interaction described above is an active file containing the JCL and selected options needed to execute the SUBDB program. When used directly to run the job, the required data subset will be stored on the requested medium in ORSER RD format, ready to be processed by any

ELK COUNTY SUBSET

NMAP (BRIGHTNESS MAP)  
OF ALL FOUR BANDS

EIGHT UNIFORM PERCENT CLASSES

Scale 1:450,000

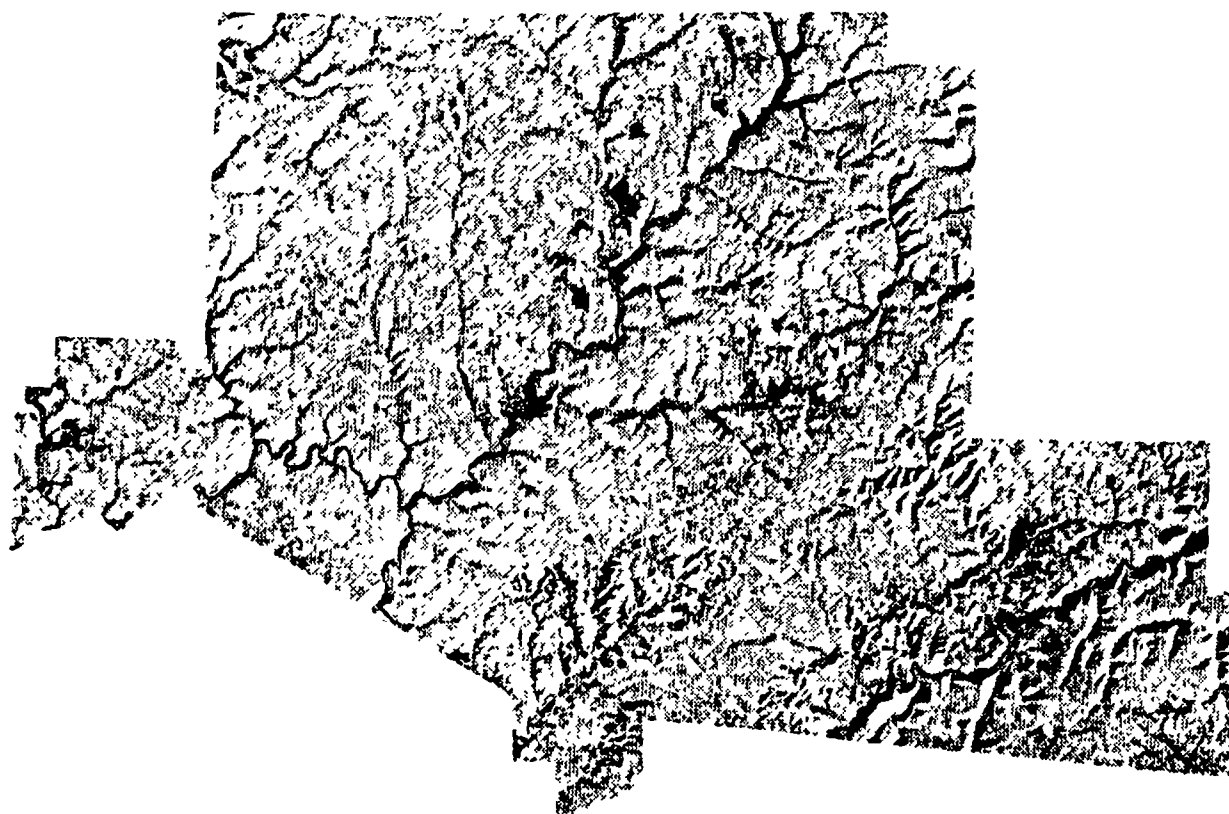


Figure 3. Electrostatic plotter print of the Elk County data set.

of the appropriate ORSER programs. (An example session with the EXEC file is given in the Appendix.)

Subset data sets are not currently cataloged within the EXEC file. At this stage, it has been sufficient to store large data sets on tapes cataloged through the ORSER tape library system (Turner et al., 1982), and to regenerate small subsets when needed. However, additional data layers, such as the recently-added binary forest/non-forest mask, may soon create a need for a cataloging system.

## 8. CONSTRUCTION OF ADDITIONAL DATA LAYERS

In addition to the forest/non-forest mask mentioned above, a Pennsylvania mosaic of summer 1981 data is nearing completion and will be registered to the data base. The western half of the mosaic (UTM 17) is being constructed at JPL, while the eastern half (UTM 18) is being constructed at ORSER. For this purpose, ORSER obtained the VICAR/IBIS software and additional mosaicking software modules from JPL, and implemented these at the Computation Center for access through the ORSER EXEC file.

The 1981 mosaic is constructed in a fashion similar to the original mosaic, except that each scene is registered to data base control points rather than to ground control points. After a week's training by a JPL representative, and the correction of some minor errors in the JPL procedures, a mosaic was produced which exactly overlaid the data base mosaic with the exception of one small area. This area was subsequently found to have too few control points. Reinstallation of some points with marginally acceptable correlations, and a repeat of the process, resulted in an exact fit.

## 9. COST ESTIMATES

The direct cost of producing a half-state (six-frame) mosaic of approximately 5250 lines and 6100 elements is approximately \$8,000. This estimate includes approximately \$3,000 for computer costs (at University rates) but excludes the cost of the data. Although this is a significant investment, such a mosaic has the advantage of being current and geographically registered to past data. In this form, subsequent processing of this data set is significantly reduced.

## 10. APPLICATIONS

The primary application of the layered mosaic is for state-wide annual assessments of defoliation of Pennsylvania forests. It is anticipated, however, that the data base will be of value to many land management and monitoring agencies throughout the state. Among the many potential applications, the following are suggested.

1. Monitoring forest resources: Much of the two-thirds of Pennsylvania covered in forest is approaching commercial maturity. Large scale changes in these forests are occurring because of harvest, mineral and fuel exploration, insect attacks, and competition from other land uses. Using the Landsat data base as the mid-date in a three-date analysis, ORSER is attempting to determine optimum change-detection procedures.



2. Soil mapping: Digitized soil maps can easily be overlaid on the data base for comparison without further rectification. The value of Landsat data for improving existing soil maps in Pennsylvania is under investigation.
3. Updating existing data bases: ORSER has developed techniques for interfacing the Landsat data base (or data derived from it) with existing geographic information systems (GIS's). The user defines a grid or polygon pattern, such as the grid-cell pattern of an existing GIS. Classified Landsat data are then extracted through this pattern and the area statistics are summarized by polygons (Irish and Myers, in preparation). Since most current land-use data bases are at the same map projection as the Landsat data base, further expensive geometric correction can be avoided.
4. Adding existing digitized information: Several types of digitized data are currently available in either raster form (e.g., digital terrain data), or in line or polygon form (e.g., roads, jurisdictional boundaries). Many of these data sets are already stored at the University Computation Center, and could easily be added as layers to the data base, if desirable.
5. Construction of small-area land cover maps: Because the significant tasks of geometric correction, and often of defining boundaries, are unnecessary when using the data base, the initial cost of these operations is spread over many projects. As a result, the cost of generating land cover maps for small geographic areas, such as watersheds and townships, is substantially reduced.

## 11. SUMMARY

The Office for Remote Sensing of Earth Resources at The Pennsylvania State University, working through a contract funded by NASA, has acquired a Landsat digital mosaic data base of the state of Pennsylvania in the UTM map projection. ORSER has also acquired the software and expertise to construct additional Pennsylvania mosaics and register them to the data base. In cooperation with personnel from the Jet Propulsion Laboratory, a state-wide summer 1981 mosaic has been constructed and registered to the data base to demonstrate the use of such data for assessment of gypsy moth defoliation. A user-friendly front-end system which permits storage, interrogation, retrieval, and manipulation of subsets of the data base and associated ancillary data, has also been developed. Thus, defoliation assessments in the state will be facilitated by the capability to quickly retrieve selected satellite imagery, and generate defoliation maps and associated statistics. In addition, the existing forest resource base map can be continually updated, enabling forest entomologists to prepare timely surveillance reports and pest management plans.

There are wide applications for the data base which, together with various ancillary data sets, can provide geographically consistent information from many sources suitable for a variety of purposes, both in research and applied fields. We anticipate that the data base will be a key source of land-use and resource data for the state.

## REFERENCES

- Cullinane Corporation. 1980. INTERACT User's Guide.
- Irish, R. R., and W. L. Myers. In preparation. An Interface for Remote Sensing Digital Image Systems and Geographic Information Systems. To be presented at the National Conference on Energy Resource Management, September 1982.
- Nelson, R. F. 1981. ASSESS2: Analysis of Four Methods for Classifying Forest Defoliation (revised). Internal Report, Earth Resources Branch, Goddard Space Flight Center.
- Turner, B. J., G. M. Baumer, and W. L. Myers. 1982. The ORSER Remote Sensing Analysis System: A User's Manual. Research Publication 109/OR. Institute for Research on Land and Water Resources, The Pennsylvania State University.
- Williams, D. L., and M. L. Stauffer. 1978. Monitoring Gypsy Moth Defoliation by Applying Change Detection Techniques to Landsat Imagery. Proceedings, Symposium on Remote Sensing for Vegetation Damage Assessment. American Society of Photogrammetry. pp. 221-229.
- Williams, D. L., M. L. Stauffer, and K. C. Leung. 1979. A Forester's Look at the Application of Image Manipulation Techniques to Multitemporal Landsat Data. Proceedings, Fifth Annual Symposium on Machine Processing of Remotely Sensed Data. pp. 368-375.
- Zobrist, A. L., and N. A. Bryant. 1979. Map Characteristics of Landsat Mosaics. Proceedings, American Society of Photogrammetry Annual Convention. pp. 260-273.
- Zobrist, A. L., N. A. Bryant, and R. G. McLeod. Unpublished manuscript. Technology for Large Digital Mosaics of Landsat Data.

? exec fro \$men.u41000.gmb.lib#orsergo on cat clr

WELCOME TO THE ORSER SYSTEM.

OK TO CLEAR ACTIVE FILE? ok

ENTER PROGRAM NAME OR 'HELP' FOR A DETAILED LIST OF INSTRUCTIONS.

ENTER 'LISTTAPES' TO LIST WORKING TAPES (RS TAPES) ASSIGNED TO YOU.

ENTER 'POLYGON' TO EXECUTE ANY ORSER POLYGON PROGRAM.

ENTER 'EXIT' TO EXIT THIS EXEC FILE.

ENTER 'DATABASE' TO ACCESS THE PENNSYLVANIA LANDSAT DATABASE.

-->database

WELCOME TO THE PENNSYLVANIA LANDSAT DATABASE.

LANDSAT DATA CAN BE RETRIEVED BY COUNTY NAME (C), BY FOREST DISTRICT (D),  
USER DEFINED POLYGON (U), OR BY PEST LOCATER GRID CELLS (P).

ENTER THE TYPE OF AREA TO BE RETRIEVED (C/D/U/P) OR TYPE 'HELP' FOR  
MORE INFORMATION.

-->c

\*\*\* THE PENNSYLVANIA LANDSAT DATABASE \*\*\*

ACCESSING AREA BY COUNTY NAME

ENTER THE COUNTY NAME. ONLY ONE COUNTY CAN BE ACCESSED AT A TIME.

ENTER 'HELP' FOR MORE INFORMATION.

-->elk

IS OUTPUT ON TAPE OR DISK? (T/D)

-->t

ENTER LAST NAME AND FIRST INITIAL SEPARATED BY ONE BLANK.

-->baumer g

ENTER OUTPUT TAPE NAME

-->rs0114

1000 COMMANDS EXECUTED WITH NO TYPING -- CHECK FOR LOOP

\* END OF COUNTY ACCESS METHOD \*

ENTER JOB PARAMETER OPTION NUMBER(S) OR 'HELP' FOR A LIST  
OF OPTIONS. TO EXIT EXEC FILE, HIT RETURN.

-->

\*\* ACTIVE FILE NOW CONTAINS STEM FOR RUNNING THE DATABASE PROGRAM \*\*  
FOR INFORMATION ON RUNNING THE PROGRAM, ENTER 'HELP', OR HIT  
RETURN TO EXIT.

-->

\*\*\* END OF ORSER EXEC FILE \*\*\*



MERGING LANDSAT DERIVED LAND COVERS INTO QUAD-REFERENCED  
GEOGRAPHIC INFORMATION SYSTEMS

James M. White  
Robert M. Ragan

University of Maryland, Remote Sensing Systems Laboratory  
College Park, Maryland 20742

and

K. Peter Lade

Department of Sociology and Anthropology, Salisbury State College  
Salisbury, Maryland 21801

There have been enough successful experiments and field applications to conclude that Landsat digital data is sufficiently accurate to define the land cover distributions required as inputs to regional planning and resource allocation models. Even though computer aided translation of raw Landsat data is extremely efficient, the adoption of this new technology by counties and other regional governments has been limited. A major problem continues to center on the difficulties of merging Landsat derived land covers into a geographic information system (GIS) that a regional government may have been using for a number of years. Typically, the data stored in such a GIS is referenced to USGS quadrangle sheets and/or state plane coordinates.

The paper describes an approach for merging multi-scene Landsat data bases into existing geographic information systems having 5-second or smaller cells. The approach uses the output from the State of Maryland's UNIVAC 1180-based Landsat classification program ASTEP (Algorithm Simulation Test and Evaluation) developed by NASA. The structure of the technique was designed to address the problems that emerged as part of the Landsat classification of the 64,000 square mile Chesapeake watershed involving twelve scenes that was conducted by the senior author as part of an EPA study. The paper describes the removal of overlap among adjacent scenes, the crossreferencing of ground control points, and the isolation of the appropriate pixels from the Landsat data base for subsequent positioning into a file containing ancillary data referenced to a specific USGS 7½ minute quadrangle sheet. Examples illustrate the clustering of classified Landsat pixels to define the dominant land use for each of 8,100 cells within a series of quadrangle sheets distributed over the State of Maryland.

The approach uses a hard copy terminal tied to an ASTEP algorithm through telephone lines. A coordinate digitizing board for inputting the position of ground control points is also valuable, although manual measurements are possible. The approach is quite efficient and should be especially attractive for use on regional scale studies.

## 1. INTRODUCTION

Many states, counties, public utilities and other organizations concerned with planning and management on a regional scale have integrated computer-based geographic information systems (GIS) into their decision making processes. Properly designed and operated systems allow the decision maker to define the spatial distribution of current conditions within the area of interest and, in an increasing number of cases, conditions in surrounding areas. Of equal importance, a good GIS also allows the decision maker to better understand how the region evolved to its current state and to interpret trends that indicate future conditions. When relatively large areas are involved, the use of GIS and computer technologies are pivotal in the development of effective planning and management strategies.

Current and past land cover distributions are key elements in the GIS. Unfortunately, these land cover files are often poorly defined or not up-to-date because of the times and costs required to assemble the data, interpret it, encode it and then enter into the GIS. Professionals concerned with GIS have long recognized the potential of digital format data from the Landsat series of satellites as a base for maintaining up-to-date land cover files in their systems. Although there are many successful applications, Landsat has remained a "potential" to the typical GIS user because the data format is less than ideal. This is especially true for grid cell based GIS that are referenced to USGS or state plane coordinate systems. Individual cells in such systems typically run from north-south vectors and may be 10, 91.8 or 4.5 acres or they may be 5 seconds in size. Although there are a number of programs designed to geometrically correct and reformat Landsat data, the time required to learn these systems, the level of effort and often the special equipment required limits the widespread application of many of the techniques and, thereby, Landsat continues to be unrealized potential.

## 2. THE CHESAPEAKE BAY EXPERIMENT

The development of the land cover distributions of the 64,000 square mile Chesapeake Bay watershed can be used to illustrate some of the problems that regional planning and management organizations encounter when attempting to integrate a Landsat derived data base into their operations. Figure 1 shows the outline of the Chesapeake watershed and the geometry of the 12 scenes used. The objectives of the Chesapeake Bay Project were: 1) produce a Level I land cover classification of the Chesapeake Bay watershed; 2) within agriculture land cover, determine tillage practices; and 3) tabulate land cover statistics by river subbasins. The land cover statistics were required as input to a mathematical model to predict the non-point source pollution loads to the Chesapeake Bay. The classification was conducted by the Northern Virginia Planning District Commission for the Environmental Protection Agency and used the IDIMS (Interactive Digital Image Manipulation System) and GES (Geographic Entry System) at NASA's Goddard Space Flight Center. The scenes had the known geometric distortions corrected (deskewing, removal of synthetic pixels) and procedures were developed to remove the overlap among scenes. The result was a properly registered land cover distribution that, through the use of a digitizer, was summarized for 63 subwatersheds distributed throughout the basin.

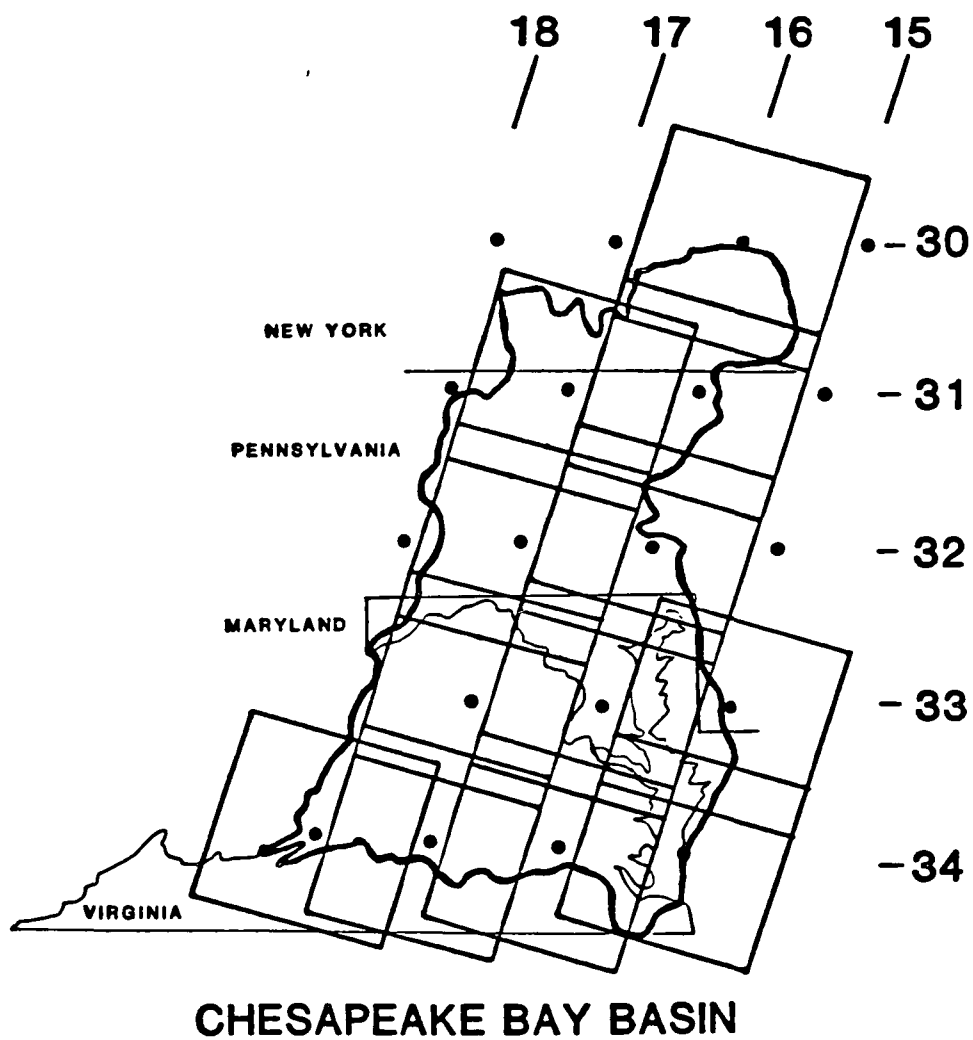


FIGURE 1

Landsat Scene Location within Chesapeake Bay Basin

The existence of such a digital data base was, obviously, very attractive to organizations within the area that had computer-based GIS. State systems, such as Maryland's MAGI (Maryland Automated Geographic Information System) and county systems, such as MSDAMP (Multi-Scale Data Analysis Mapping Program) used by Montgomery County, Maryland appeared to be the logical recipients of the data derived in the Chesapeake study. While it was straight forward to extract a relatively large polygon such as a watershed from the Chesapeake Landsat data base, MAGI and MSDAMP require the definition of land covers within individual cells referenced to USGS or state plane coordinates. The general concept of MSDAMP is illustrated in Figure 2. MSDAMP is a series of 90 x 90 five second cells referenced to 19 USGS 7½ minute quadrangle sheets. The user obtains information by entering the name of the quadrangle sheet or sheets of interest and then extracts information by defining either polygons or individual cells. Maryland's MAGI uses either a 91.8 or 4.54 acre cell. Geographical Information Systems of the MSDAMP and MAGI types must have one dominant land cover defined for each cell in the data base. A schematic of the definition problem is illustrated in Figure 3. The domain of a particular USGS quadrangle sheet must be isolated from the Landsat data base and then a specific five second cell must become computer retrievable to the staff of the user organization. Image processing capabilities are available to all Maryland state and local governmental organizations through the State's UNIVAC 1100 series of computers located at the University and State College campuses. As potential State and county users of the Maryland portion of the Chesapeake data base moved toward integrating this additional information into their GIS, it became obvious that the efforts were not going to be widely successful because the needed software did not exist in a form that was compatible with UNIVAC 1100 series computers. There was no parallel software that could: rotate the Landsat coordinate system; reference the individual cells to USGS coordinates; isolate an array of cells defining a USGS 7½ minute quadrangle sheet and then resample the individual pixels to define a single land cover category for a predefined cell size. Without such software, the Chesapeake data base provided an excellent source of qualitative information, but remained inaccessible to the day-to-day user of the established computer-based geographical information systems operating within the State.

### 3. OBJECTIVES

If the Chesapeake Landsat-derived data base and similar future Landsat efforts are to be integrated into the existing geographical information systems, it is necessary to develop additional software to overcome the problems discussed above. To be useable, the additional software has to be fully integrated into established computer based approaches that are accessible and familiar to the users. Because few of the users in the State of Maryland have access to color CRT-based interactive image processing systems, the software had to be designed to run on a standard UNIVAC 1108 mainframe computer and require no more than a modem-connected hard copy terminal for operation. Further, because of severe restrictions placed on core storage during the daytime hours, the system had to be designed for minimum core storage utilization. With these constraints in mind, system development was undertaken to meet the following objectives:

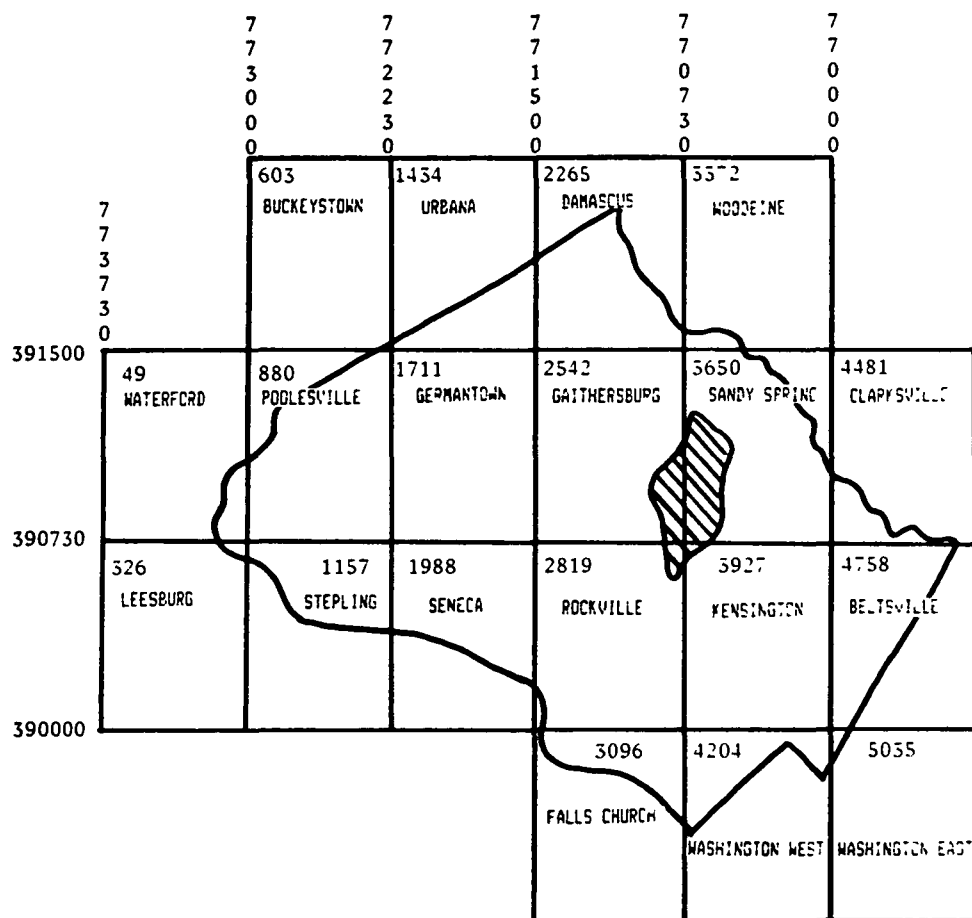


Figure 2

Quad-Sheet Storage Arrangement For  
Montgomery County, Maryland

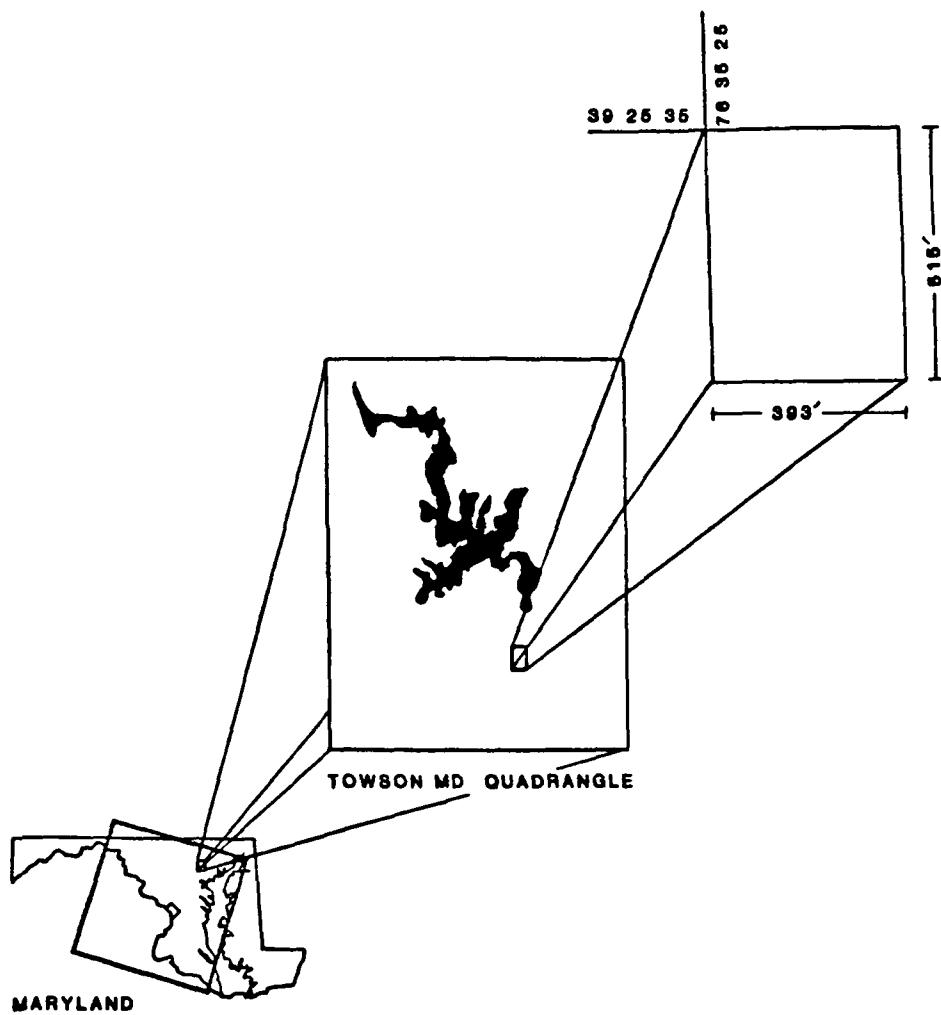


Figure 3  
Isolation of a Single Cell from a Landsat Scene

- 1) Develop interactively an equation which relates a Landsat coordinate system to a latitude longitude coordinate system.
- 2) Create a transformed data base from Landsat imagery compatible with preconfigured Geographical Information Systems.
- 3) Enter geographic data from a map surface into ASTEP.

#### 4. SYSTEM CAPABILITIES

To meet the objectives listed above, two programs (REGISTER and TRANSFORM) were developed. The program REGISTER was designed to input geographic data from a map surface and develop a regression relating the Landsat coordinate system to a latitude, longitude coordinate system. The program TRANSFORM, using the equations developed by the program REGISTER, was designed to create a geometrically correct data base compatible with preconfigured geographical information systems.

The program REGISTER serves two functions. First, it outputs to a file the longitude and latitude of points digitized from a map surface. Second, it provides an equation relating the latitude, longitude of a point to its Landsat line and sample coordinate. To complete the first function, a link (equation) must be developed relating the position on a map surface to its latitude and longitude. The position on the map surface may be input as coordinates from a digitizing table or measured manually off the map using the upper left corner as the origin. (Note: While manually measuring the location of points on a topographic map may be tedious, if it is done carefully, accuracy on a 1:24000 scale map can be  $\pm 40$  feet.) The user inputs to the program are the longitude and latitude in degrees.minutes.seconds of the upper left corner of the map, the size of the map in minutes, the distance between "tic-marks" on the map in minutes, and the coordinates of the "tic-marks" from the digitizing table or as measured manually by the user. A first order polynomial regression equation is then developed relating the coordinate from the map surface to its longitude and latitude coordinate. A list of the actual and predicted coordinates, as well as the residuals, is produced for each of the "tic-marks" are output. The user has the option of removing any of the "tic-marks" from the registration if they were incorrectly digitized. The user also has the option of changing the regression equation to second or third order. (Note: For large scale maps, i.e., 1:24000 there should be no need to go to a second or third order equation.) Once the map has been registered to the digitizing table, the location of the ground control points can be digitized from the map. These points are then stored in a file for use in developing the transformation equation.

The second function that the program REGISTER performs is to allow the user to develop an equation relating the latitude, longitude coordinate system of the Geographical Information System to the line and sample coordinate of Landsat. The program reads the file containing the ground control points created above and a least-square file is applied to the points to develop a simple linear transformation of the form:

$$\hat{X} = C(1) + C(2)Y + C(3)X$$

$$\hat{Y} = C(9) + C(10)Y + C(11)X$$

where  $\hat{X}$  and  $\hat{Y}$  are the estimated sample and line values in the Landsat coordinate system, X and Y are the observed values of longitude and latitude (digitized coordinates) in the GIS coordinate system, and C(1), C(2)...C(N) are the coefficients for the transformation equation expressed in the form:

$$[\hat{Y}\hat{X}] = [1YX] \begin{matrix} C(1) & C(9) \\ C(2) & C(10) \\ C(3) & C(11) \end{matrix}$$

An output table is printed that contains the estimated sample and line value, the observed sample and line value, and the error (observed-estimated) sample and line value for each ground control point.

Upon examination of the ground control points, the user has the option of altering the list of ground control points. The user is prompted: DO YOU WISH TO EDIT POINTS? Y/N. If the user responds with an upper case Y he is prompted with: ADD(A) DELETE(D) OR EXIT(E)?. If the user wishes to delete a point, he responds with an upper case D. (Note: the development of the equation is an iterative process, the user may wish to restore a ground control point that was previously deleted by responding A.) The user is then prompted: INPUT NUMBER(S) TO BE ADDED OR DELETED ZERO (0) TO END. The user then would input the number(s) of the ground control point to be deleted, 0 indicates there are no more points. The user is then prompted: ADD(A) DELETE(D) OR EXIT (E)? and would respond E. The first order regression equation is then recalculated and the output table is again listed. The user has the option of editing points and recalculating the first order regression equation until he is satisfied that all the ground control point residuals have the same order of magnitude. When the prompts to edit points are answered N, the user will be prompted with: DO YOU WISH THIS TO BE THE HIGHEST ORDER? Y/N. If the response is N, a second order equation is developed with the form:

$$\hat{X} = C(1) + C(2)Y + C(3)X + C(4)Y^2 + C(5)X^2 + C(6)XY$$

$$\hat{Y} = C(9) + C(10)Y + C(11)X + C(12)Y^2 + C(13)X^2 + C(14)XY$$

The output table is printed and the user is given the option of editing points or developing a third order equation. The third order equation has the form:

$$\hat{X} = C(1) + C(2)Y + C(3)X + C(4)Y^2 + C(5)X^2 + C(6)XY + C(7)Y^3 + C(8)X^3$$

$$\hat{Y} = C(9) + C(10)Y + C(11)X + C(12)Y^2 + C(13)X^2 + C(14)XY + C(15)Y^3 + C(16)X^3$$

When the final transformation equation has been calculated, the equation is stored in a disc file for use by the program TRANSFORM.



The program TRANSFORM reads the transformation equation developed above and prompts the user for information concerning the location and size of the transformed area. The program prompts the user with: INPUT LONGITUDE, LATITUDE OF UPPER LEFT CORNER OF STUDY AREA IN DD.MMSS. When the user responds, he is prompted: INPUT THE SIZE OF TRANSFORMED AREA IN MINUTES LONGITUDE, LATITUDE. The user is not constrained to having the size of the transformed area the same in both longitude and latitude. The user is then prompted for the number of cells in the X and Y directions in the transformed area. The program TRANSFORM displays the Landsat sample and line value for the four corners of the transformed area and the minimum subset of the original data needed to transform the area. (Note: This allows the user to redefine the original study area to use the minimum amount of computer storage and CPU time.) The user then has the option of stopping the run to subset the original data, or continuing the run and creating the transformed area.

Figure 4 is a schematic representation of the procedure to transform a study and form a Landsat line and sample coordinate system to a latitude, longitude coordinate system. The output from the program TRANSFORM is a file that can be read directly into a geographic information system or reformatted by a program (INASTEP) to be entered back into ASTEP to use its statistical and map generating capabilities.

## 5. PROCEDURE

The procedure to transform a study area from a Landsat referenced coordinate system into a georeferenced coordinate system is as follows:

- 1) Output lineprint maps of study area
- 2) Locate and digitize features that can be found on both lineprint maps and topographic maps.
- 3) Develop regression equation
- 4) Transform the data.

The first step in transforming the data is to output lineprint maps from ASTEP, such as that illustrated in Figure 5, of the study area for use in locating features (ground control points). The lineprint map generation is the most critical portion of locating ground control points. A lineprint map is limited to displaying one channel of data with a practical limit of 20 grey levels, therefore, whatever a user can do to combine information from more than one MSS channel of data on a lineprint map is important. There are many ASTEP output products which are useful in the production of lineprint maps. A grey level map (density slice) of channel 7 can provide good land/water interface detection; it can also be useful in locating bridges, river boundaries, and power line clear cuts. A grey level map of channel 5 is useful in finding man-made features such as road intersections and industrial parks.

There are three ASTEP routines that allow the user to output information from more than one MSS channel of data. A map from the norm of all four channels (brightness map) can often be used to augment the output products

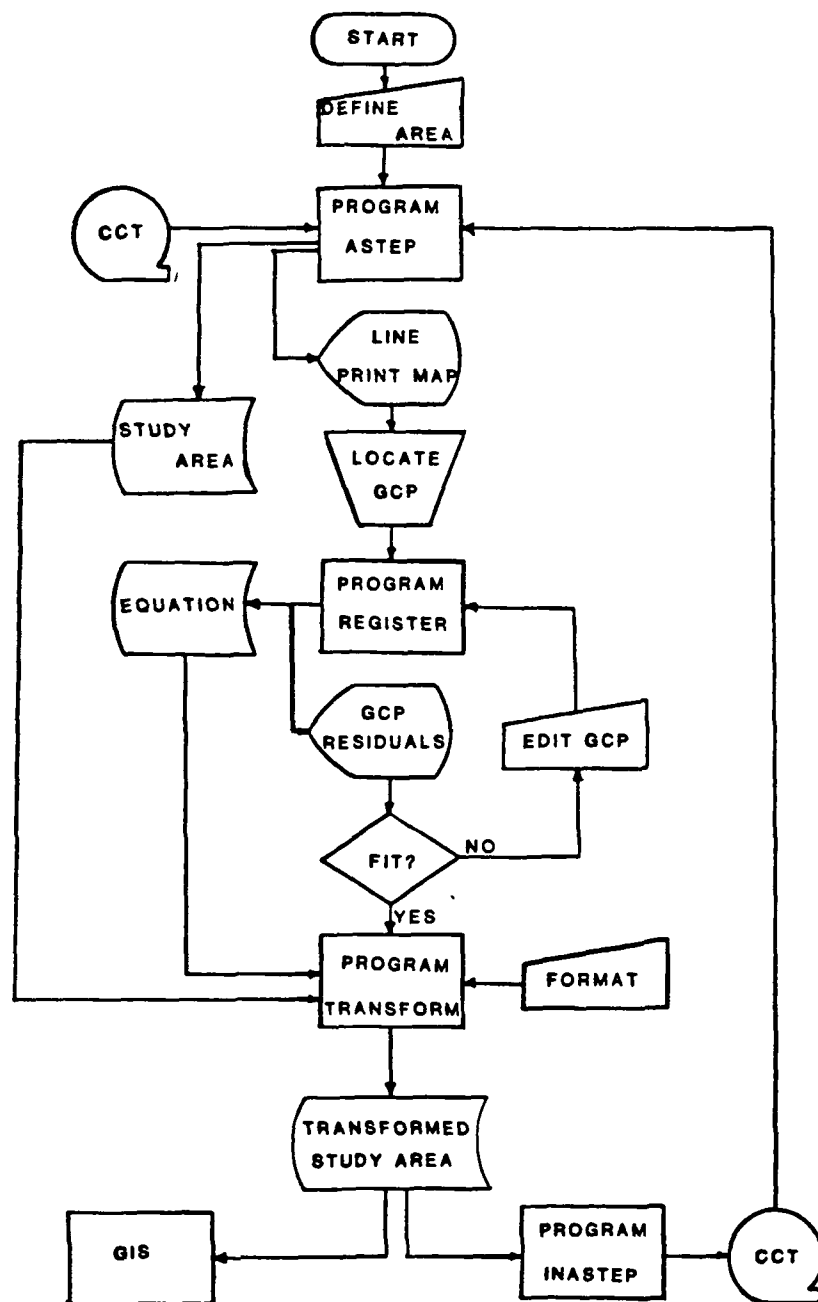


Figure 4  
Procedure to Transform Landsat Data



FIGURE 5  
Unsupervised Classification of Study Area

listed above. An unsupervised classification of the study area using relatively few classes (6-8) provides a quick method of locating forest and grass field boundaries. Figure 5, for example, is an example of an unsupervised classification using only 6 classes.

Once the lineprint maps of the study area have been generated, the ground control point location can begin. It is important to find ground control points that are uniformly distributed throughout and surrounding the study area. A general rule of thumb is that in order to have confidence in the coefficients, there should be at least four ground control points for each of the coefficients of the regression equation. Therefore, a first order equation should have a minimum of 12 ground control points, a second order should have a minimum of 24 ground control points, and a third order should have a minimum of 32 ground control points. Depending on the size of the study area, the land cover, and the topography, it may not be feasible to find as many as 30 ground control points using lineprint maps. The ground control point can be any fixed feature locatable on both the lineprint maps and topographic maps. They may include bridges, islands, road intersections, power line clear cuts, and small ponds. It is generally easier to locate ground control point from images in early spring or late fall when there are no leaves on the trees to obscure ground features. There are some features however, that are easier to locate in summer scenes (i.e., power lines, roads).

When all the ground control points have been located, two files are created for each topographic map in the study area. The first file contains the digitized coordinates of the "tic-marks" on the topographic map. The second file contains the digitized coordinates and Landsat sample and line coordinates for each of the ground control points.

After the ground control points have been digitized, the process of developing the regression equation can begin. Tables I - V are examples from a program runstream which illustrates the process of developing a regression equation. For the sake of simplicity, only control points from one topographic map will be used. The user responses are underlined and comments are in brackets.

TABLE I

Initial Output Used to Verify Regression Equation Defining  
Coordinates of Points on Quadrangle Sheet

EXQT RSSL\*REGISTER.ABSTS

INPUT 1 TO DIGITIZE,2 TO DEVELOP REGRESSION,0 TO QUIT

1

ENTER LONGITUDE AND LATITUDE OF THE UPPER LEFT CORNER  
SIZE OF TOPO IN MINUTS(LON,LAT),DISTANCE BETWEEN THE  
REGISTRATION POINTS IN MINUTS(LON,LAT)

76.3730 39.3000 7.5 7.5 2.5 2.5

INPUT REGISTRATION POINTS

ADD TRY1. [TRY1. is the file that contains the digitized coordinates of  
the "tic-marks"]

#	X/	Y/	X	Y	EX	EY
1	76.37300	39.30001	76.37300	39.30000	-.0000004	-.0000084
2	76.35001	39.30000	76.35000	39.30000	-.0000082	-.0000002
3	76.32300	39.30000	76.32300	39.30000	-.0000059	.0000004
4	76.30001	39.30000	76.30000	39.30000	-.0000086	-.0000012
5	76.37299	39.27299	76.37300	39.27300	.0000121	.0000064
6	76.34599	39.27299	76.35000	39.27300	.0000066	.0000070
7	76.32300	39.27300	76.32300	39.27300	.0000016	.0000035
8	76.30000	39.27300	76.30000	39.27300	-.0000012	.0000000
9	76.37300	39.24600	76.37300	39.25000	.0000039	.0000018
10	76.34599	39.25000	76.35000	39.25000	.0000066	-.0000018
11	76.32298	39.25000	76.32300	39.25000	.0000164	.0000006
12	76.29599	39.25000	76.30000	39.25000	.0000113	-.0000010
13	76.37302	39.22300	76.37300	39.22300	-.0000219	-.0000051
14	76.35000	39.22299	76.35000	39.22300	-.0000016	.0000053
15	76.32301	39.22301	76.32300	39.22300	-.0000094	-.0000141
16	76.29600	39.22299	76.30000	39.22300	.0000008	.0000082

#	X COEFF	Y COEFF
1	276326.42283975470368000000	141730.89427985332736000000
2	.01966168255291389490	19.76698345728208928000
3	-25.46664236535343654400	-.00409724894996088550

ERROR SQ = .18980407714843749984

SUM ERR X= .023438 SUM ERR Y = .015625

DO YOU WISH TO EDIT POINTS? Y/N

N

DO YOU WISH THIS TO BE THE HIGHEST ORDER? Y/N

Y

INPUT GROUND CONTROL POINTS # THEN DIGITIZED COORDINATES AND LABEL

ADD TRY. [TRY. contains the digitized coordinates of the ground control  
points and their line and sample coordinates.]

EOF

INPUT 1 TO DIGITIZE,2 TO DEVELOP REGRESSION,0 TO QUIT

Table I is a list of the predicted and actual longitude and latitude of the "tic-marks" on the topographic map, as well as the errors (actual-predicted).

where:

$X_1$  = predicted longitude of "tic-mark"

$Y_1$  = predicted latitude of "tic-mark"

$X$  = actual longitude of "tic-mark"

$Y$  = actual latitude of "tick-mark"

$EX = X - X_1$

$EY = Y - Y_1$

In this example, all the errors are less than 0.2 of a second (approximately 16' at this latitude) so there was no need to edit points or increase the order of the regression equation. After the user replies Y to the prompt "DO YOU WISH THIS TO BE THE HIGHEST ORDER?" the user is prompted for the ground control points and digitized coordinates.

The program uses the equations generated in Table I to convert the digitized coordinates of the ground control points to their corresponding longitude, latitude and stores the results for later use. The process is repeated for each topographic map in the study area. After all maps have been digitized, the user responds "2" to the prompt "INPUT 1 TO DIGITIZE, 2 TO DEVELOP REGRESSION, 0 TO QUIT".

TABLE II

Output Used to Verify Regression Equation

2

#	X/	Y/	X	Y	EX	EY
1	1598.32462	125.30842	1600.00000	126.00000	1.6753845	.6915751
2	1700.71982	121.36526	1707.00000	121.00000	6.2801819	-.3652592
3	1752.95477	102.56207	1767.00000	103.00000	14.0452271	.4379263
4	1673.43864	162.95376	1670.00000	164.00000	-3.4386444	1.0462418
5	1766.56570	142.39142	1772.00000	143.00000	5.4342957	.6085815
6	1800.19225	214.15857	1794.00000	212.00000	-6.1922455	-2.1585655
7	1749.11740	170.16807	1749.00000	170.00000	-.1174011	-.1680679
8	1811.80373	223.93128	1804.00000	223.00000	-7.8037262	-.9312840
9	1793.64261	169.85357	1795.00000	169.00000	1.3573914	-.8535690
10	1693.76706	247.05843	1674.00000	245.00000	-19.7670593	-2.0584335
11	1730.37199	283.31932	1775.00000	290.00000	44.6280060	6.6806755
12	1799.00040	254.53150	1783.00000	252.00000	-16.0003967	-2.5314999
13	1764.99077	209.58345	1758.00000	209.00000	-6.9907684	-.5834541
14	1718.71126	179.44728	1714.00000	180.00000	-4.7112579	.5527229
15	1678.39890	195.36758	1670.00000	195.00000	-8.3988953	-.3675785
#	X COEFF		Y COEFF			
1	133544.14964234083856000000		38163.43098602816448000000			
2	-.21203176960660829952		-.38703929184634677568			
3	-.36905888861332414208		.06160473314834291592			
ERROR SQ =		3226.59675617842002880000				
SUM ERR X =		.000092	SUM ERR Y =		.000011	

The program reads the file containing the latitude, longitude and line sample values for each ground control point and develops a regression equation relating latitude, longitude to line and sample. Table II is a list of the actual line and sample, predicted line and sample and the errors for each of the ground control points. Where:

$X_1$  = predict sample value

$Y_1$  = predict line value

$X$  = actual sample value

$Y$  = actual line value

$EX = X - X_1$

$EY = Y - Y_1$

The user is prompted "DO YOU WISH TO EDIT POINTS? Y/N". Deciding which point(s) to remove from the regression equation is somewhat of an art. A good rule-of-thumb would be to remove any point whose errors are significantly different from the rest (i.e., point 11). Being an iterative process, the user can delete points to see the effects and later add them if he wishes. In this example, the point "11" is deleted.

TABLE III

Output Used to Verify Regression Equation Without Point #11

DO YOU WISH TO EDIT POINTS? Y/N

Y

ADD(A) DELETE(D) OR EXIT(E) ?

D

INPUT NUMBER(S) TO BE ADDED OR DELETED ZERO(0) TO END

11

0

ADD(A) DELETE(D) OR EXIT(E) ?

E

#	X/	Y/	X	Y	EX	EY
1	1597.25696	125.22393	1600.00000	126.00000	2.7430420	.7760677
2	1706.68803	122.19216	1707.00000	121.00000	.3119659	-1.1921606
3	1765.75558	104.30714	1767.00000	103.00000	1.2444153	-1.3071394
4	1669.70222	162.44365	1670.00000	164.00000	.2977753	1.5563526
5	1772.50732	143.16842	1772.00000	143.00000	-.5073242	-.1684227
6	1794.33221	213.26280	1794.00000	212.00000	-.3322144	-1.2628002
7	1748.62469	170.05817	1749.00000	170.00000	.3753052	-.0581665
8	1804.76753	222.86302	1804.00000	223.00000	-.7675323	.1369820
9	1795.93893	170.10218	1795.00000	169.00000	-.9389343	-1.1021824
10	1675.03110	244.43800	1674.00000	245.00000	-1.0310974	.5619984
12	1785.26913	252.53663	1783.00000	252.00000	-2.2691345	-.5366306
13	1757.85748	208.53445	1758.00000	209.00000	.1425171	.4655514
14	1714.56311	178.84788	1714.00000	180.00000	-.5631104	1.1521187
15	1668.70555	194.02156	1670.00000	195.00000	1.2944489	.9784431

#	X COEFF	Y COEFF
1	134280.80787729471872000000	38173.89989193249484800000
2	-.15026574780891621632	-.37825928840648126272
3	-.40357207301249786560	.05704062922797348256

ERROR SQ = 31.58291904046200204800

SUM ERR X = .000122 SUM ERR Y = .000011

Table III lists the output for the regression equation developed without ground control point "11". There are no ground control points having errors significantly different from the rest, so the user responds N to the prompt "DO YOU WISH TO EDIT POINTS? Y/N". The user is then prompted with "DO YOU WISH THIS TO BE THE HIGHEST ORDER? Y/N". If the user is not satisfied with the size of the errors, he will respond "N" and a second order equation will be developed, as illustrated in Table IV.

TABLE IV

Output to Verify Regression Equation Using Second Order Equation

DO YOU WISH TO EDIT POINTS? Y/N

N

DO YOU WISH THIS TO BE THE HIGHEST ORDER? Y/N

N

#	X/	Y/	X	Y	EX	EY
1	1599.11740	126.62065	1600.00000	126.00000	.8825989	-.6206522
2	1707.99907	122.20280	1707.00000	121.00000	-.9990692	-1.2027960
3	1767.02563	103.49714	1767.00000	103.00000	-.0256348	-.4971447
4	1670.40628	163.15906	1670.00000	164.00000	-.4062805	.8409386
5	1772.93092	142.48143	1772.00000	143.00000	-.9309235	-.5185738
6	1793.28535	212.64375	1794.00000	212.00000	.7146454	-.6437511
7	1748.69167	169.81061	1749.00000	170.00000	.3083344	.1893864
8	1803.47786	222.15224	1804.00000	223.00000	.5221405	.8477650
9	1795.66638	169.25774	1795.00000	169.00000	-.6663818	-.2577400
10	1674.15500	245.56256	1674.00000	245.00000	-.1549988	-.5625591
12	1783.61646	252.21816	1783.00000	252.00000	-.6164551	-.2181587
13	1757.14552	208.36650	1758.00000	209.00000	.8544769	.6335011
14	1714.68178	179.08262	1714.00000	180.00000	-.6817780	.9173832
15	1668.79240	194.93917	1670.00000	195.00000	1.2075958	.0608253

#	X COEFF	Y COEFF
1	55196.07290533184998400000	8554.40744912996884480000
2	-.73650654382436186816	.15635408307161924240
3	.46676819108142808384	-.00889545607060426843
4	.00000245849135216414	-.00000080964223510283
5	-.00000147942547933447	.00000041454940314387
6	-.00000037657992861539	-.00000110603872397266

ERROR SQ = 13.09050019155256448000  
SUM ERR X = .008270 SUM ERR Y = .005571

DO YOU WISH TO EDIT POINTS? Y/N

N

DO YOU WISH THIS TO BE THE HIGHEST ORDER? Y/N

Y

END PROGRAM REGISTER.



The output shown in Table IV is again listed for the second order equation and the user has the option to edit points or go on to a third order equation. Once the user responds "N" to each question, the program creates a file with the coefficients of the regression equation.

After the regression equation has been developed, the user can transform the study area. The program TRANSFORM using the regression equation developed above, prompts the user for location and format of the transformed area. The program reads the raw Landsat data from a file created by ASTEP, transforms the data and outputs the transformed data in a format compatible with various geographic information systems. The user will be prompted for the longitude, latitude of the upper left corner of the study area, and the size of study area in minutes. The user is not required to have the study area correspond to one topographic map, and the study area can have different dimensions in the latitude and longitude direction. Table V is an example run of the program TRANSFORM; the study area is the Towson, MD quadrangle. (See Figure 6) The output file is to have the data stored in 5 second cells.

TABLE V

Example Row for Towson, MD

INPUT LON,LAT OF UPPER LEFT CORNER OF AREA IN D.MS

76.3730 39.3000

INPUT SIZE OF STUDY AREA IN MINUTS LON,LAT

7.5 7.5

INPUT NUMBER OF CELLS LON,LAT IN TRANSFORMED AREA

90 90

LANDSAT COORDINATES OF TOPO SHEET

1590., 121.\*\*\*\*\*1770., 93.

I  
I  
I  
I  
I

I  
I  
I  
I  
I

1655., 292.\*\*\*\*\*1834., 264.

IF YOU WISH TO SUBSET	STARTING LINE	STARTING SAMPLE#	LINES #	SAMPLES
	88	1584	209	256

DO YOU WISH TO SUBSET ? Y/N

Y

DO YOU WISH TO QUIT ? Y/N

Y

END PROGRAM TRANSFORM



FIGURE 6

USGS 7,5' Topographic Map Towson, MD

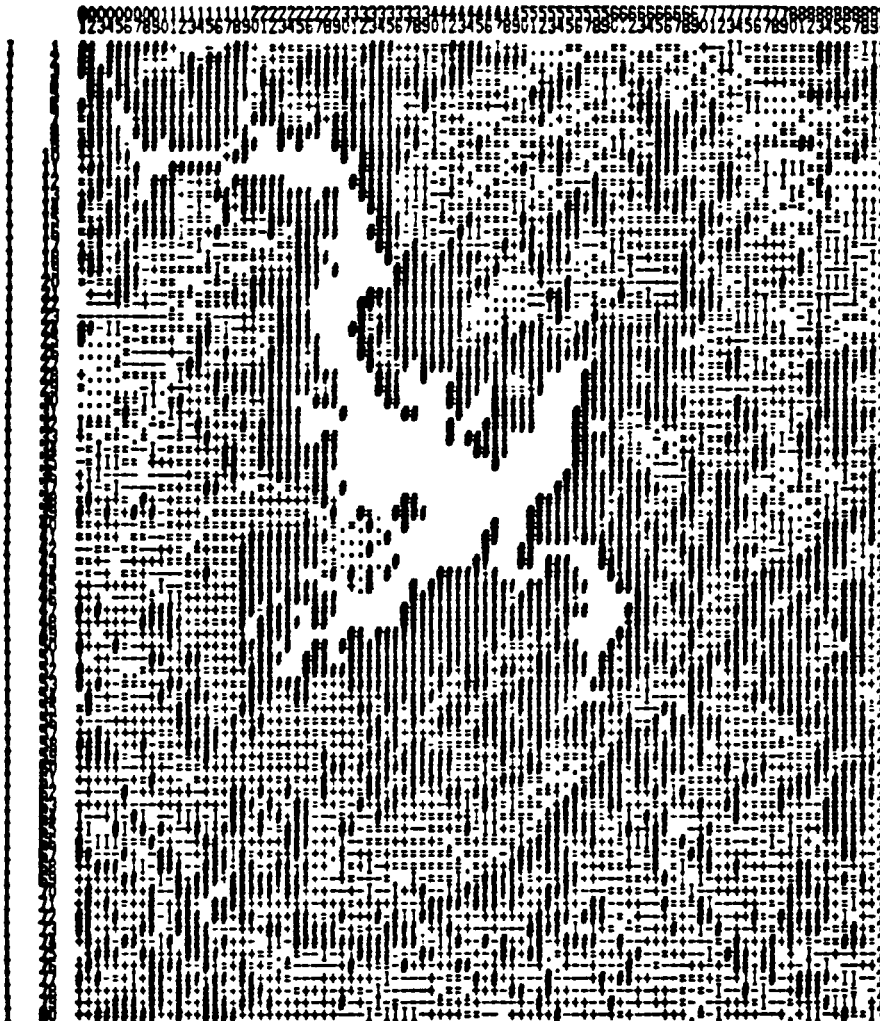
Figure 5 is the raw Landsat data used in the example in Table V, and Figure 7 is a map of the output file.

The user, by changing the number of cells in the output transformed area can produce output products at a given scale (i.e., 1:24000). If in the example above, had changed the number of cells in the transformed area from 90 x 90 to 212 x 272, a map of the output file would have a scale of 1:24000. Figure 8 is an example of such a product.

## 6. CONCLUSION

Many current or potential users of digital format remotely sensed imagery are restricted to the use of a remote lineprinter type terminal that accesses processing software on a general purpose, mainframe computer. The software described in the present paper was designed to provide this group of users with some of the interactive geometric corrections and data manipulation capabilities found on dedicated, color CRT-based image processing systems such as IDIMS. The system developed is compatible with ASTEP input/output routines and the UNIVAC 1100 series core limitations. It requires only a typewriter type terminal and is, therefore, available to Maryland State and local government users.

The interactive editing capabilities allow the user to produce a  $\pm 1$  pixel registration accuracy between an image and map referenced position. Flexible output format routines allow interfacing with preconfigured geographical information systems. With minor modifications, the system can easily be adapted to other geographical formats (i.e., state plane, UTM) and other sensors (i.e., RBV). The resulting transformed data bases can be re-entered into the ASTEP program to allow the user access to ASTEP capabilities such as scaled map production and statistical tabulations.

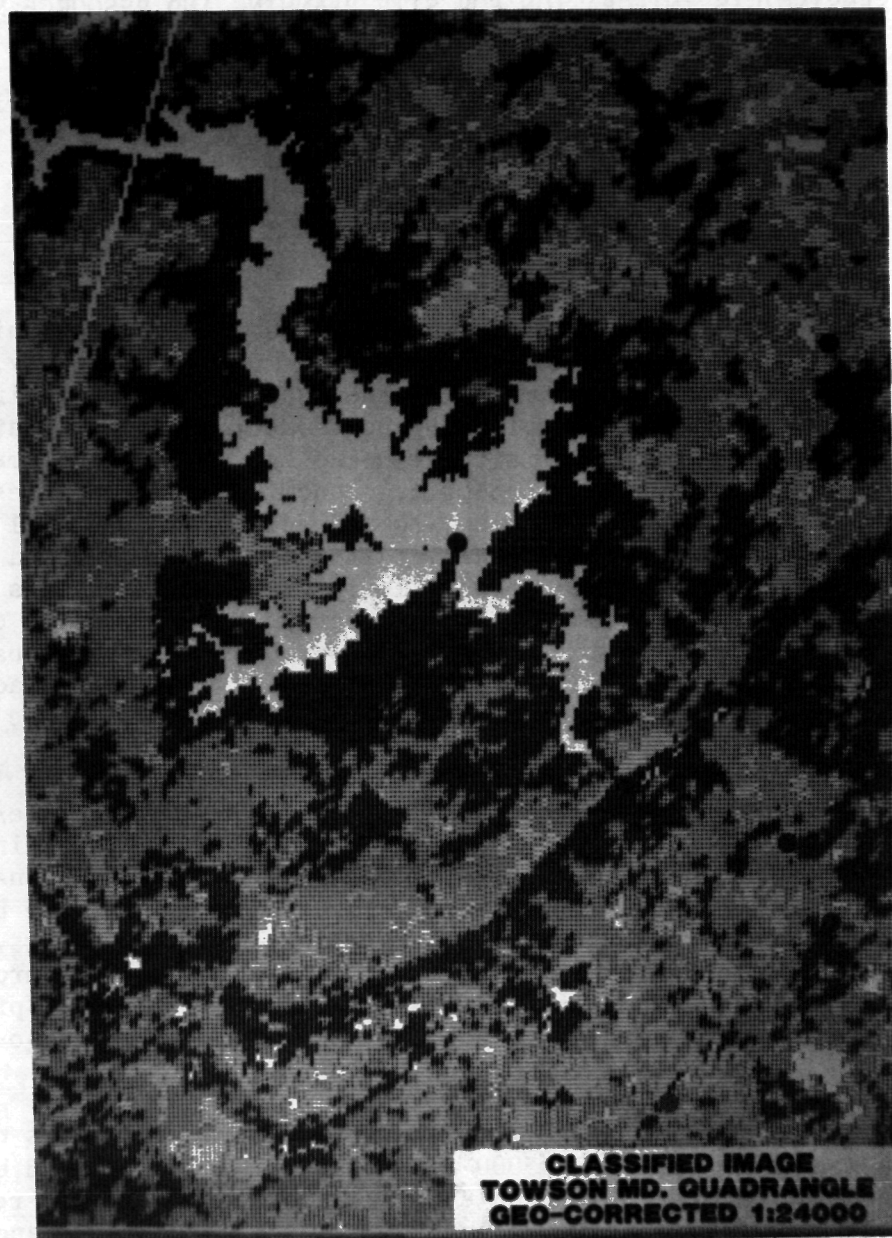


**CLASSIFIED IMAGE**

**TOWSON MD. QUADRANGLE**

**GEO-CORRECTED 90x90 5 SECOND CELLS**

Figure 7



**CLASSIFIED IMAGE  
TOWSON MD. QUADRANGLE  
GEO-CORRECTED 1:24000**

FIGURE 8

Quad-Centered Transformation Scale 1:24000

## REMOTE SENSING/GIS INTEGRATION FOR SITE PLANNING AND RESOURCE MANAGEMENT

J. D. Fellows

University of Maryland, Department of Civil Engineering  
College Park, Maryland 20742

-

### ABSTRACT

In the late 1970's, the Maryland National Capital Park and Planning Commission, the county level planning offices for Montgomery County, Maryland, was faced with the problems of managing the rapid growth of their 400 square mile jurisdiction. The planning board decided digital data bases would be constructed capable of site planning and defining parameters necessary for energy, planning, hydrologic, economic and various resource forecasting models. Techniques for managing and collecting regional ancillary data (aerial photos, soils, etc.) and digital remote sensed data sets would be required to run models and produce timely graphics/statistics for present/proposed decision-making. Concurrently, the University of Maryland Civil Engineering Department (UOMCE) was developing a computer-based gridded information system (GIS) allowing engineers to create, access, integrate, and maintain a multi-parameter geographical data base for real-time hydrologic modeling.

By joining forces, the UOMCE and MNCPPC developed an interactive/batch GIS (array of cells georeferenced to USGS quad sheets) and interfacing application programs (e.g., hydrologic models). This system allows non-programmer users to request any data set(s) stored in the MNCPPC data base by inputting any random polygon's (watershed, political zone) boundary points. The data base information contained within this polygon can be used to produce maps, statistics, and define model parameters for the area. Present/proposed conditions for the area may be compared by inputting future usage (land cover, soils, slope, etc.).

This system, known as the Hydrologic Analysis Program (HAP), is currently operational on the MNCPPC's HP 3000 mini-computer and the UOMCE UNIVAC-1180 main-frame computer. HAP has been especially effective in the real-time analysis of proposed land cover changes on runoff hydrographs and graphics/statistics resource inventories of random study area/watersheds.

## INTRODUCTION

In August of 1977, the Montgomery County Offices of The Maryland National Capital Park and Planning Commission incorporated an adaptation of a computer-based geographical information system known as MSDAMP. MSDAMP, an acronym for Multi-Spatial Data Analysis Mapping Program, was developed at the Iowa State University Land Analysis Laboratory in November 1972. This effort allowed cultural and physical data from existing maps and aerial photographs to be converted to a digital format and stored on the County's computer. The dominant land use or other desired parameter for each of these cells is encoded and entered into the computer. In batch mode, MSDAMP can be used to produce line printer gray-scale maps showing the distribution of various land uses, geologic features, slopes, soils, etc.

The data structure of MSDAMP is an array of five-second cells covering the entire 400 square mile Montgomery County. Each cell, encompassing an area of 4.58 acres, measures 397.75 feet east/west and 505.90 feet north/south. These dimensions were selected to give the five second increment at a latitude of 39°00'00" and allow distortion-free symbolic maps to be produced with ten column by eight line high speed line printers. At this cell resolution, each data plane (land cover, soils, etc.) in the Montgomery County data base would contain 55895 cells. MSDAMP requires the input of each individual cell georeferenced by latitude and longitude. In the subsequent years MSDAMP proved to be generally ineffective because of this cumbersome data collection method, inability to interface with application programs (e.g., hydrologic and planning models) and inefficient sequential data base searches.

In 1978, the University of Maryland Civil Engineering Department (UOMCE) was conducting research in geographical information systems (GIS) for hydrologic analysis. The goal of the UOMCE GIS was to improve large area geoencoding techniques, provide a data structure capable of manipulating random polygons of data within the data base, interface the data base with existing hydrologic models, and manage the data base consisting of both digital remote sensed data (Landsat, digital terrain) and ancillary hard copy information (aerial photography and soil maps).

By joining forces, the MNCPPC and UOMCE created a system called the Hydrologic Analysis Program (HAP) which satisfied both organizations goals. The MNCPPC involvement ensured that the HAP software/hardware requirements would be developed for a non-programmer county level production environment.

It was decided that the HAP data collection, data management, and model interface would be tested by expanding the MNCPPC data base to include land cover, hydrologic soil groups and slope necessary for defining parameters in the Soil Conservation Service hydrologic model, TR-55. In a typical real-time situation, the planner types the coordinates of a watershed boundary on the keyboard of an office terminal connected with a Hewlett-Packard 3000 mini-computer. The appropriate HAP compatible MSDAMP

data files are automatically accessed and the planner's terminal immediately produces:

1. a series of symbolic maps showing the distribution of land use, soil type, and slope;
2. statistical tables showing the number of acres and the percentage of the watershed area devoted to each land use, soil type, or slope category;
3. a table showing the runoff curve number, hydraulic length, and time of concentration needed to enter the SCS model.

If the planner also enters a rainfall amount into the terminal, HAP will automatically list the volume of runoff and estimated peak discharge as computed with SCS-TR-55.

A major function of HAP is to give planners the capability to assess the impact of land use or other changes being considered for the watershed. Thus, after the above information has been obtained for existing conditions, the planner can type the coordinates of those subareas being considered for change into the keyboard. He lists the type of new land use or other changes being considered for each of these subareas. The program then recycles in the manner outlined above with the new subarea information to give data for the watershed under new conditions.

#### HAP MODEL

The guiding principle in the development of HAP was that planners must be able to obtain quantitative information in real time for any watershed/political zone in the County. This pilot HAP model consists of three major components: 1) an enhanced GIS designed for collecting and manipulating the data base; 2) the interface between the data base and SCS-TR-55 hydrologic model; and 3) graphics/statistic package.

#### HAP Gridded Information System

The HAP GIS was designed to simplify the user's maneuvering of data between the hierarchical levels of the data base. This task was accomplished by providing the user with a hierarchical STUDY AREA-7½° UNITED STATES GEOLOGICAL SURVEY (USGS) TOPO SHEET-GRID-CELL-ATTRIBUTE model. Figure 1 shows the Montgomery County, Maryland study area subdivided into its nineteen USGS topo sheets. Each topo consists of 90x90 grid of 4.58 acre cells containing various geophysical attributes. The cells are geo-referenced by latitude/longitude and grid row/col.

The concept of storing geographical information in a computer can be illustrated by Figure 2. Conceptually, the spatial distribution of the geographical quantities are coded as an array of rectangular cells with the position of each cell identified by a row and column number. The dominant



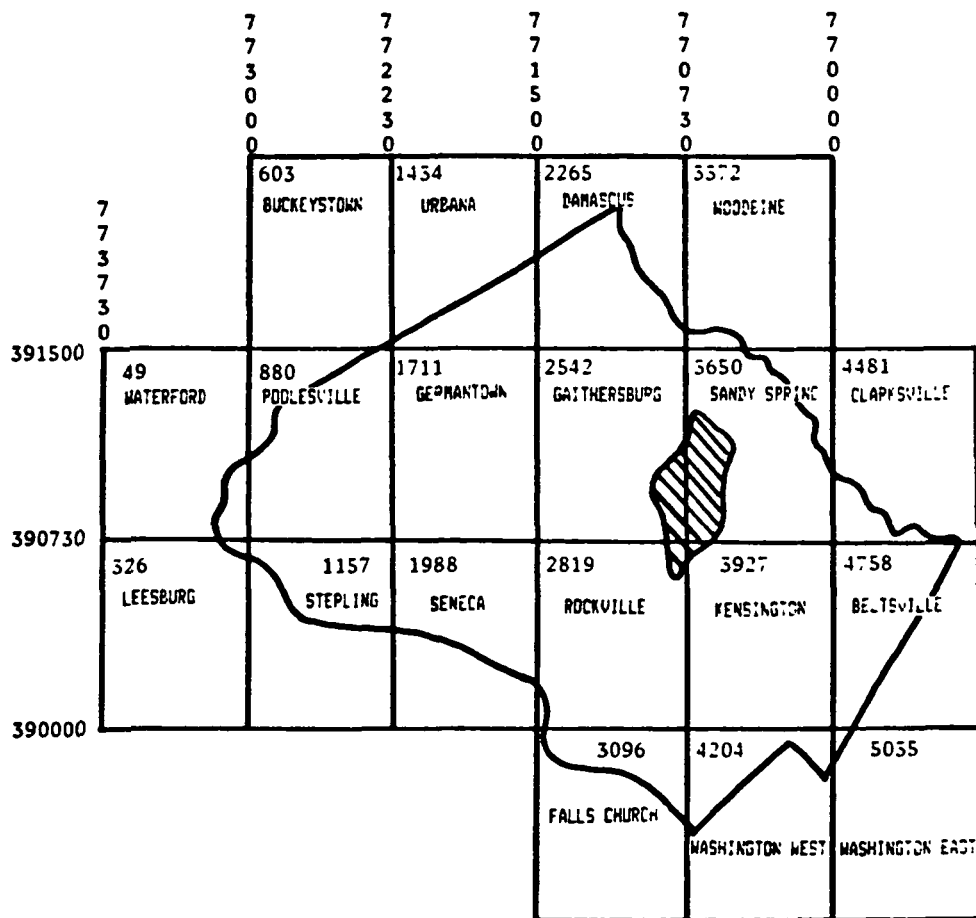


Figure 1  
Quad-Sheet Storage Arrangement For  
Montgomery County, Maryland

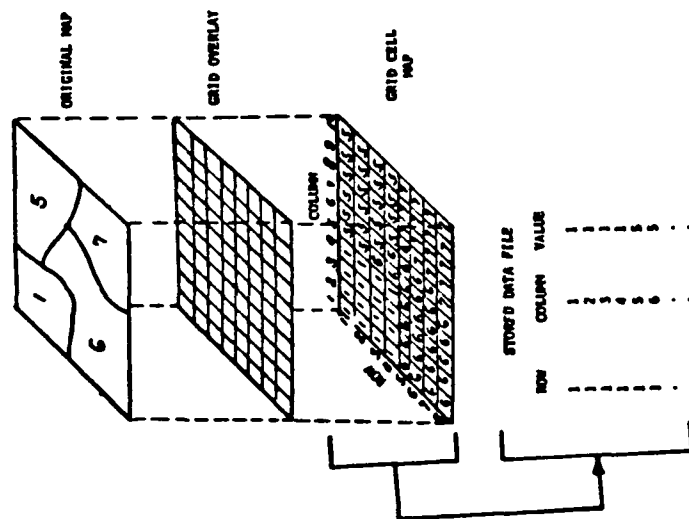


Figure 2  
Concept of Single Variable  
Data Bank  
(from the Corps of Engineers<sup>1)</sup>)

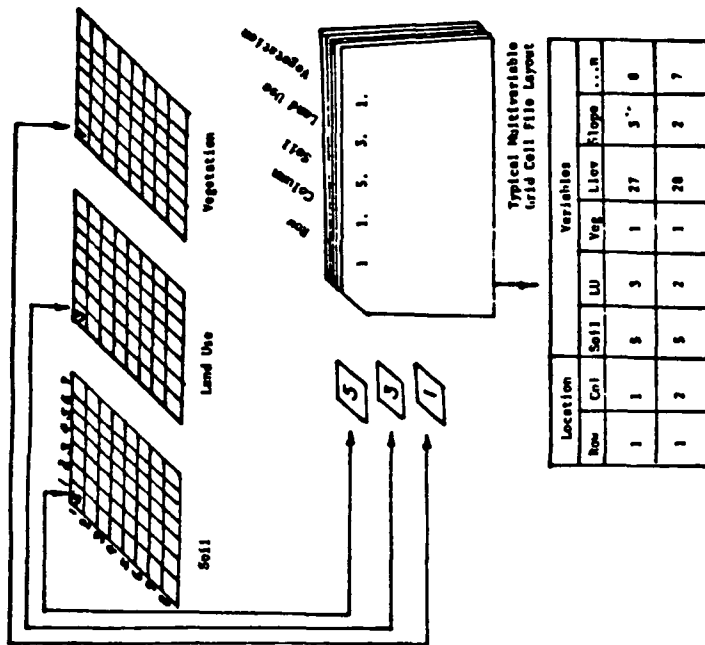


Figure 3  
Storage of Multivariable Information  
In A Grid Cell Data Base  
(from the Corps of Engineers<sup>1)</sup>)

geographical quantity within each cell is identified, generally, by a single valued alphanumeric character. This information is stored in the computer in some format similar to that listed in the lower part of Figure 2. The strategy illustrated by Figure 2 can be extended to a system in which layers of geographic information are reduced to the same format, added to the computer and interfaced with each other to produce multi-variant parameters, such as those required for hydrologic models. Figure 3 is a schematic illustrating the organization of a multi-parameter data base.

Using the HAP model as a loading template, HAP will accept data by the grid-overlay featured in Figure 2, digitized polygons or previously existing gridded digital data. The electronic digitization of single-valued polygons is very efficient for homogeneous data attributes like hydrologic soil groups. The polygons are converted to grid cells within HAP. To minimize collection efforts, an entire topos grid is tagged with the dominant data class. Only the cells containing subsequent data classes are actually geocoded.

The model also allows an engineer or planner to use the quad sheet as a "pointer" and, thereby, manipulate only the data within the quad sheets involved in an analysis, rather than the entire MSDAMP data base. For example, if a planner was interested in the analysis of the shaded watershed shown in Figure 1, the first step would be to type in the topo access numbers 2542, 2819, 3650, and 3927. This would cause the computer to access the tape or other off line inexpensive storage and bring the 90 x 90, 5-second arrays of cells contained within the Gaithersburg, Sandy Spring, Rockville, and Kensington quadrangle sheets into temporary direct access storage. The role of the topo access will be explained in a subsequent section.

Table I lists the land use categories accessed by HAP from the MSDAMP data base. Unless specified by the planner from the terminal, the symbols listed in Table I will be used by HAP in printing any maps requested in the output. The planner has an array of options including the assignment of "blanks" to some of the categories, or all but one, in order to produce special purpose thematic maps. The Curve Numbers listed in Table I will be discussed in the following section.

Table II lists the slope categories and their map symbols stored in the HAP data base. Ranges of slopes, rather than specific slope values, are used in order to allow the slope within a cell to be represented as a single digit and, thereby, minimize the storage requirements.

#### HAP Hydrologic Model

The power of the system is realized when these arrays of stored variables are interfaced directly through the computer with simulation models. In this approach, cells within a watershed are combined to define the input parameters needed for the hydrologic model. The model then outputs desired streamflow characteristics in a format appropriate to the user's requirements.

Table I

Land Cover Symbols and Curve Numbers  
Used In Hydrologic Analysis Program

LAND COVER CATEGORY	SYMBOL	CURVE NUMBER FOR SOIL GROUP			
		A	B	C	D
Grass	B	39.	61.	74.	80.
Cultivated Fields	C	72.	81.	86.	91.
Conifer Forests	E	25.	55.	70.	77.
Deciduous Forests	F	25.	55.	70.	77.
Idle Lands	G	49.	69.	79.	84.
Rural Residential	J	51.	68.	79.	84.
Industrial/commercial	K	89.	92.	94.	95.
Single Family	L	61.	75.	83.	87.
Low Density	M	57.	72.	81.	86.
High Density	N	77.	85.	90.	92.

Table II

Slope Category Symbols Used In Hydrologic Analysis Program

SLOPE RANGE	SYMBOL
0.0 LE* PERCENT SLOPE LT* .25	A
.25 LE PERCENT SLOPE LT .50	B
.50 LE PERCENT SLOPE LT .75	C
.75 LE PERCENT SLOPE LT 1.0	D
1.0 LE PERCENT SLOPE LT 1.5	E
1.5 LE PERCENT SLOPE LT 2.0	F
2.0 LE PERCENT SLOPE LT 2.5	G
2.5 LE PERCENT SLOPE LT 3.0	H
3.0 LE PERCENT SLOPE LT 4.0	I
4.0 LE PERCENT SLOPE LT 6.0	J
6.0 LE PERCENT SLOPE LT 8.0	K
8.0 LE PERCENT SLOPE LT 10.0	L
10.0 LE PERCENT SLOPE LT 12.5	M
12.5 LE PERCENT SLOPE LT 15.0	N
15.0 LE PERCENT SLOPE LT 20.	O
20.0 LE PERCENT SLOPE LT 25.0	P
25.0 LE PERCENT SLOPE LT 30.0	Q
30.0 LE PERCENT SLOPE LT 40.0	R
40.0 LE PERCENT SLOPE LT 50.0	S
50.0 LE PERCENT SLOPE LT 75.0	T
75.0 LE PERCENT SLOPE LT 100.	U
100. LE PERCENT SLOPE LT 200.	V
200. LE PERCENT SLOPE LT 300.	W
300. LE PERCENT SLOPE LT 500.	X
500. LE PERCENT SLOPE	Y

LE - LESS OR EQUAL TO

\*LT - Less than

The model used to generate the volumes of runoff and peak discharges from the information available in the data base is a computerized version of parts of SCS-TR-55<sup>2</sup>. It is assumed that the HAP user has or will develop an understanding of SCS-TR-55. Thus, only the key equations needed to explain the logic flow of HAP will be presented.

The implementation of HAP starts with the user drawing a boundary around the watershed of interest on USGS 7½ minute quad sheets. Names of the quad sheets involved and a sufficient number of points to adequately define the boundary are entered into the terminal. The internal software of HAP then isolates the cells contained within the watershed boundary. The land use and hydrologic soil group are interfaced to compute a curve number for each cell in accordance with Table I. The cells are then summed and divided by the total to obtain an average Curve Number for the watershed. In a similar fashion, the average slope is obtained for the watershed. Table III lists the symbols that will be used to print HAP soil maps if desired.

After the Curve Number, CN, has been derived, the relationship

$$S = \left( \frac{1000}{CN} \right) - 10 \quad (1)$$

is then used to compute the potential maximum storage, S, of the watershed. The result of this computation is then entered into

$$Q = \frac{(P - .2S)^2}{P + .8S} \quad (2)$$

to obtain the volume of runoff, Q, from the rainfall, P.

HAP obtains the area of the watershed, A, by counting the number of cells encompassed within the boundary and multiplying by 4.58. The area is then available to estimate the hydraulic length in accordance with

$$H_L = 209(A)^{.6} \quad (3)$$

The time of concentration is then estimated from

$$T_c = \frac{1}{0.6} \left( \frac{H_L^{.8} (S+1)^{.7}}{1900 Y^{.5}} \right) \quad (4)$$

where Y is the average slope of the watershed. Finally, the time of concentration is entered into a mathematical relationship defining the curve of Figure 4, to produce a dimensionless peak discharge which HAP converts to the peak discharge in cubic feet per second by multiplying the area in sq. miles by the volume of runoff obtained from Equation 2.

Table III

SOIL TYPE SYMBOLS USED IN HYDROLOGIC ANALYSIS PROGRAM

Group	Symbol
A	A
B	B
C	C
D	D

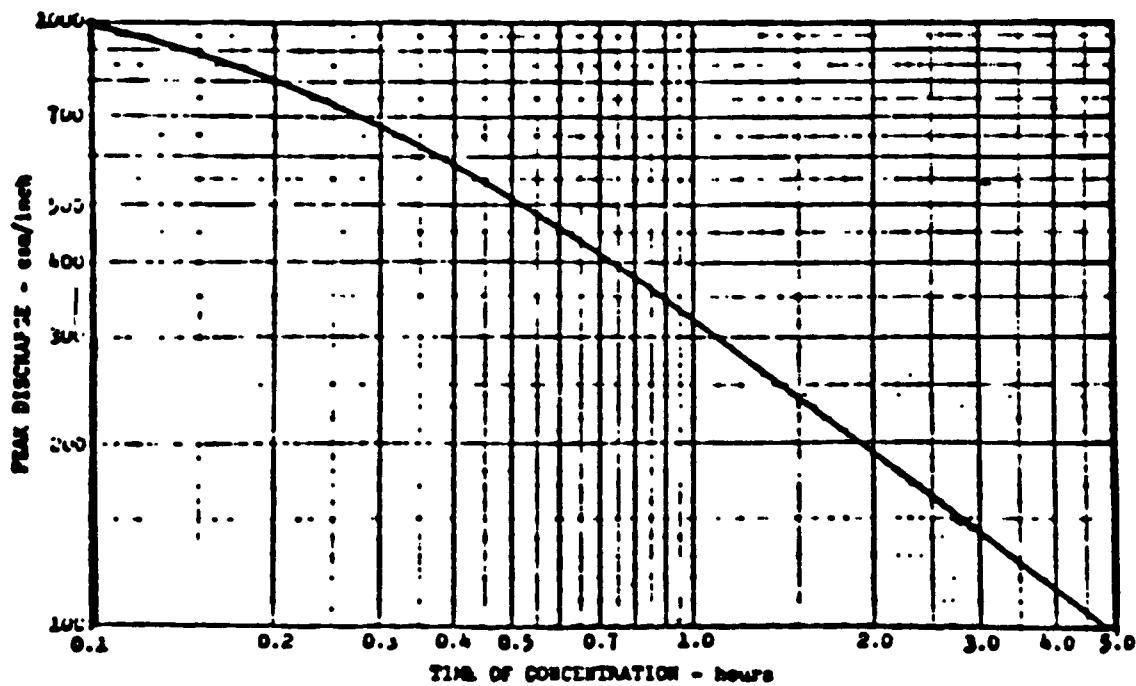


Figure 4

Peak Discharge in scin Per Inch of Runoff Versus Time of Concentration ( $T_C$ ) for 24-hour, Type-II Storm Distribution.

The use of Figure 4 imposes the requirement that the rainfall used with HAP be a 24 hour volume so it will be consistent with the intensity distribution of the SCS Type-II Storm Rainfall. The SCS Type-II Storm is excellent for use in urban and urban fringe studies. The storm structure provides a period of several hours of light rainfall during which the soils are "wetted-up" to reduce infiltration rates and fill a portion of the depression storage. There is then a period of several hours of heavy rainfall simulating thunderstorm intensities. Finally, there is a period of lightening rainfall during the last several hours.

It should be recognized that there are a number of options within SCS-TR-55. One approach within TR-55 is to use detailed maps to measure incremental hydraulic flow lengths and then develop the time of concentration through velocity computations. HAP defines the hydraulic length and the time of concentration with the empirical Equations 3 and 4 which were developed by SCS from regression analyses of watersheds throughout the United States. A constraint on Equation 4 is that it be limited to watersheds of less than 2000 acres. Table IV contains Montgomery County rainfall-frequency values required by TR-55.

Table IV

24 Hour Rainfall in Inches for Different Frequency Events  
for Montgomery County

<u>Frequency</u>	<u>Rainfall Inches</u>
1	2.6
2	3.2
5	4.2
10	5.1
25	5.6
50	6.3
100	7.2

## USE OF THE HYDROLOGIC ANALYSIS PROGRAM FOR WATERSHED INVESTIGATIONS

The use of HAP as a tool for watershed investigations will be illustrated through the use of an example problem. The specific problem addressed is the comparison of the peak discharge for a watershed under present conditions with the peak discharge anticipated under conditions of complete development as allowed by the Montgomery County approved land use master plan. The statement of the example problem is as follows:

There is a need to provide stormwater storage behind an existing conduit on the Rockville quad. The design of the control structure that will create the backwater requires estimates of the peak discharge for present and future land use conditions for a 100 year, 7.2 inch, 24 hour storm. Determine the current and future peak discharges, volumes of runoff and percent changes created in these quantities by developing the watershed to its ultimate land use distribution.

HAP uses four "modules" during its execution. The function of these modules are:

**BOUNDARY MODULE=** The BOUNDARY module is called when the user wishes to enter the coordinates of the watershed boundary. The BOUNDARY module retrieves and stores the cells contained within the boundary of the study area.

**HYDROLOGIC MODEL=** The HYDROLOGIC MODEL module merges the cells within the boundary and computes the slope for the study area. These parameters are then interfaced with the SCS-TR-55 and volumes of runoff and peak discharges estimated for user input rainfalls. The user is also prompted for map production and statistical summaries of the study area for present conditions (i.e., land cover, soils, and slope).

**UPDATE MODULE=** The UPDATE module allows the input of proposed cell changes within the study area. The cell distribution in the watershed is adjusted for the proposed conditions interfaced with TR-55 and revised runoffs compared to the present conditions. The user is prompted for production of maps and statistical summaries describing the study area under proposed conditions.

**STOP MODULE=** The STOP module is called when the user has completed his study and wishes to exit from the HAP program.



The solution of the example problem will assume the user is accessing the program capabilities with a 132 character terminal. The program is interactive with the terminal outputting prompts and explanations to aid the user in inputting data or requesting HAP operations. Those terminal statements within boxes are input statements entered by the user. Those statements in upper case letters are outputs from the terminal.

The following run stream describing the steps in the solution of the example problem contains explanations of key points.

#### TYPICAL HYDROLOGIC ANALYSIS PROGRAM RUN STREAM

-----

#### STEP ONE            GATHER NEEDED INFORMATION PRIOR TO ACCESSING HAP

Before working with the terminal, it is necessary, in the interest of efficiency, to tabulate the information necessary to define the watershed boundary and move the appropriate portion of the data base into direct access storage. The first step is to assemble the topographic sheets needed to define the watershed boundary. The watershed boundary is sketched. The 90 x 90 transparent mylar grid is overlaid on the quad sheet. Enough nodes are picked to allow the watershed boundary to be approximated as a series of straight line segments. In this example, the entire watershed is within the bounds of the Rockville quadrangle sheet. In order to get the data of this quadrangle into direct access storage, the user will have to enter the disk access number obtained from Figure 1, in this case 2819. The user will also have to enter the latitude and longitude of the upper left hand corner of the quadrangle sheet. Finally, the cells to be changed to a proposed land cover will have to be tabulated for entry from the terminal. Thus, prior to accessing the terminal, the user would write down the information below.

Latitude = 390730  
Longitude = 771500  
Disk Access No. = 2819

Tabulation of cell coordination of watershed boundary. Proceed clockwise. The boundary will automatically be closed between the first and last points listed.

Watershed Boundary

Boundary Point	Row	Column
1	04	50
2	04	53
3	06	55
4	08	55
5	09	56
6	11	56
7	12	55
8	14	55
9	15	56
10	17	56
11	18	55
12	19	54
13	18	52
14	17	51
15	15	51
16	15	49
17	14	48
18	12	48
19	11	47
20	07	47

Tabulation of cells to be changed to reflect proposed watershed conditions. The cell location and proposed land cover change within each cell is tabulated using the transparent mylar grid and the land cover codes of Table I . If soil type or slope were to also be changed, Tables II and III would be used to obtain the proper symbols.

LOCATION		NUMBER OF CELLS	LAND COVER	SOILS	SLOPE
ROW	COLUMNS				
4	50-52	3	k	zero	zero
5	49-54	6	k	"	"
6	50-55	6	k	"	"
7	49-55	7	k	"	"
8	49-55	7	k	"	"
9	49-56	8	k	"	"
10	48-56	9	k	"	"
11	47-56	10	k	"	"
12	48-53	6	k	"	"
13	50-53	4	k	"	"
14	50-53	4	k	"	"
15	51-52	2	k	"	"

If more than one quadrangle is involved, this process will be repeated for each sheet.

-----  
STEP TWO

SIGN ON HEWLETT PACKARD 3000 AS FOLLOWS  
-----

- (a) Turn on Terminal
- (b) Wait for green light
- (c) Hit return key
- (d) You will get a colon back
- (e) Type in (HELLO USER.MPLENV,WATER)
- (f) It will ask you for Password
- (g) Type in ENVIRON
- (h) Type in NAZ
- (i) You will get # sign
- (j) To execute HAP you type in HAP

THE TERMINAL INPUTS AND OUTPUTS FOR STEP 2 ARE:

\*\*\*\*\*HYDROLOGIC ANALYSIS PROGRAM\*\*\*\*\*  
\*\*ENTER: BOUNDARY, HYDROLOGIC MODEL, UPDATE, LOAD, STOP\*\*

Note: At this point the terminal will stop. The output immediately above lists the modules available. HAP is ready to accept the input of the watershed boundary.

-----  
STEP THREE

BOUNDARY MODULE  
-----

- A. Request the BOUNDARY module by typing BOUNDARY into the terminal.
- B. Specify if you need input prompts.
- C. Enter name of study area.
- D. Enter number of topos containing study area.
- E. Enter latitude and longitude of northwest topo corner.
- F. Enter DAF topo address (see Figure 1)
- G. Enter whether boundary points are P=POLYGON or C=CELL format.
- H. Enter whether boundary points are I=INTEGER or D=DIGITAL.
- I. In this example the boundary is defined from the northeast corner row and column coordinates of a 90 x 90 (4.58 acre) cell grid. If the boundary points were entered from a digitizer the map scale would be requested. The program is set up to accept the node points from a digitizer in inches (eight sets of X, Y coordinates per card in F5.2 format).

THE TERMINAL INPUTS AND OUTPUTS FOR STEP 3 ARE:

```

BOUNDARY
*****BOUNDARY MODULE*****

**DO YOU NEED DATA INPUT PROMPTS(Y/N)**
Y
**NAME OF STUDY AREA (MAX 30 CHAR)**
ROCKVILLE WATERSHED ANALYSIS
**ENTER: NUMBER OF TOPOS CONTAINING STUDY AREA**
11
**ENTER: TOPO LATITUDE, LONGITUDE, BIGN ADDRESS.
(*P*=POLYGON OR "C" CELLS DATA SET), (*I*=INTEGER OR *D*=DIGITAL DATA)
390730,771500,2819,2,1
**ENTER: ROW, COL (TERMINATE WITH ENDO)**
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651

```

\*\*\*\*\***MOBILE MOBILE**\*\*\*\*\*

\*\*INPUT RAINFALL IN INCHES (REAL NUMBER) \*\*

74

cc-WATERSHED ANALYSIS (RCS-TN-23)cc

IDENTIFICATION-ROCKVILLE WATERSHED ANALYSIS

WEIGHTED CURVE NUMBER= 71.75  
AVE PERCENT SLOPE= 3.67  
AREA= 531.01 (ACRES) .03 (SQUARE MILES)

HYDRAULIC LENGTH= 8028.18 (FEET)  
TIME CONCENTRATION= 2.05 (HOURS)  
RAINFALL= 7.20 (INCHES)  
VOLUME OF RUNOFF= 3.97 (INCHES)  
PEAK DISCHARGE= 838.83 (CFS)

IN ACCORDANCE WITH THE SCS METHOD OF ESTIMATING VOLUME OF  
RUNOFF THE TIME OF CONCENTRATION AND HYDRAULIC LENGTH ARE  
4.010 FOR WATERSHEDS OF LESS THAN 2000 ACRES (SCS-TP-149)

SHOULD YOU LIKE A MAP OF PRESENT CONDITIONS(Y/N)??

04

.....MAP NOBLE.....

++MAP REQUIRES A TERMINAL WITH MINIMUM 80 CHARACTER WIDTH++  
++ENTER OUTPUT UNIT++



5-DO YOU WANT A MAP OF MONT. CO RES. PHOTO LAND COVER (Y/N)??



DO YOU WANT DEFAULT OR NEW MAP SYMBOLS (B/N)?=

1-3

[illegible][illegible]

280715	4	OFFK	4	280715
280710	5	OFFICE	5	280710
280705	6	KAFFBMC	6	280705
280700	7	KAFFBMC	7	280700
280655	8	KAFFBMC	8	280655
280650	9	KAFFBMC	9	280650
280645	10	CCCCCB888	10	280645
280640	11	CCCCCB888	11	280640
280635	12	CCFFB88	12	280635
280630	13	KAFFFFF	13	280630
280625	14	KAFFFFF	14	280625
280620	15	KAFFFFF	15	280620
280615	16	KAFFFFF	16	280615
280610	17	KAFFFFF	17	280610
280605	18	KAFFFFF	18	280605
280600	19	KAFFFFF	19	280600
280555	20	KAFFFFF	20	280555
280550	21	KAFFFFF	21	280550
280545	22	KAFFFFF	22	280545
280540	23	KAFFFFF	23	280540
280535	24	KAFFFFF	24	280535
280530	25	KAFFFFF	25	280530
280525	26	KAFFFFF	26	280525
280520	27	KAFFFFF	27	280520
280515	28	KAFFFFF	28	280515
280510	29	KAFFFFF	29	280510
280505	30	KAFFFFF	30	280505
280500	31	KAFFFFF	31	280500
280455	32	KAFFFFF	32	280455
280450	33	KAFFFFF	33	280450
280445	34	KAFFFFF	34	280445
280440	35	KAFFFFF	35	280440
280435	36	KAFFFFF	36	280435
280430	37	KAFFFFF	37	280430
280425	38	KAFFFFF	38	280425
280420	39	KAFFFFF	39	280420
280415	40	KAFFFFF	40	280415
280410	41	KAFFFFF	41	280410
280405	42	KAFFFFF	42	280405
280400	43	KAFFFFF	43	280400
280395	44	KAFFFFF	44	280395
280390	45	KAFFFFF	45	280390
280385	46	KAFFFFF	46	280385
280380	47	KAFFFFF	47	280380
280375	48	KAFFFFF	48	280375
280370	49	KAFFFFF	49	280370
280365	50	KAFFFFF	50	280365
280360	51	KAFFFFF	51	280360
280355	52	KAFFFFF	52	280355
280350	53	KAFFFFF	53	280350
280345	54	KAFFFFF	54	280345
280340	55	KAFFFFF	55	280340
280335	56	KAFFFFF	56	280335
280330	57	KAFFFFF	57	280330
280325	58	KAFFFFF	58	280325
280320	59	KAFFFFF	59	280320
280315	60	KAFFFFF	60	280315
280310	61	KAFFFFF	61	280310
280305	62	KAFFFFF	62	280305
280300	63	KAFFFFF	63	280300
280295	64	KAFFFFF	64	280295
280290	65	KAFFFFF	65	280290
280285	66	KAFFFFF	66	280285
280280	67	KAFFFFF	67	280280
280275	68	KAFFFFF	68	280275
280270	69	KAFFFFF	69	280270
280265	70	KAFFFFF	70	280265
280260	71	KAFFFFF	71	280260
280255	72	KAFFFFF	72	280255
280250	73	KAFFFFF	73	280250
280245	74	KAFFFFF	74	280245
280240	75	KAFFFFF	75	280240
280235	76	KAFFFFF	76	280235

[illegible]





-----  
**STEP FIVE            ENTER PROPOSED CHANGES OF STUDY AREA AND COMPARE SCS TR-55  
 RESPONSE FOR PRESENT/PROPOSED CONDITIONS**  
 -----

The heading listing the modules has output on the terminal. The SCS-TR-55 analysis for current conditions is complete. HAP is ready to accept changes within the watershed and repeat the SCS-TR-55 analysis. If there are no watershed changes to be investigated, proceed to Step Six to exit program.

- A. Request the UPDATE module.
- B. As prompted, select the quads requiring changes.
- C. Enter the location and proposed land cover, soil, or slope value for each update cell from Tables I, II, and III. In this case, only land covers are being changed.
- D. For the given rainfall in STEP FOUR, the updated conditions will be routed through TR-55 and the response compared with the present conditions.
- E. The user will be prompted for proposed condition maps to verify the update input.

THE TERMINAL INPUTS AND OUTPUTS FOR STEP 5 ARE:

```

UPDATE
=====UPDATE MODULE=====
ARE THERE ANY CHANGES IN THE 390730 771500 TOPO(Y/N)**
Y
**ENTER: ROW,COL,OCUR,LAND COVER,SOIL,SLOPE (EX: 10,30,4,A,B,C)
**IF NO CHANGE INPUT A ZERO (TERMINATE WITH ENDO)**
4,30,3,K,0.0
5,49,6,K,0.0
6,50,6,K,0.0
7,49,7,K,0.0
8,49,8,K,0.0
9,49,8,K,0.0
10,48,9,K,0.0
11,47,10,K,0.0
12,48,6,K,0.0
13,30,4,K,0.0
14,30,4,K,0.0
15,31,2,K,0.0
ENDO

==WATERSHED ANALYSIS (SCS-TR-55)===

IDENTIFICATION=ROCKVILLE WATERSHED ANALYSIS
RAINFALL(INCHES)= 7.20
DRAINAGE AREA(ACRES)= 531.81 .03 (SQUARE MILES)
HYDRAULIC LENGTH(FEET)= 8028.18

QUANTITY                      PRESENT      PROPOSED
CURVE NUMBER                  71.75      86.23
AVERAGE PERCENT SLOPE        3.67      3.67
TIME OF CONCENTRATION(HOURS) 2.05      1.30
VOLUME OF RUNOFF              9.87      9.98
PEAK DISCHARGE(CFS)          839.83    1244.69

***CHANGE IN RUNOFF            1.61 (INCHES)    48.37 %***
***CHANGE IN PEAK DISCHARGE    405.06 (CFS)     48.60 %***

*****
IN ACCORDANCE WITH THE SCS METHOD OF ESTIMATING VOLUME OF
RUNOFF THE TIME OF CONCENTRATION AND HYDRAULIC LENGTH ARE
VALID FOR WATERSHEDS OF LESS THAN 2000 ACRES (SCS-TP-149)
*****
  
```





• •

• •

4

**D**

[illegible][illegible]

四

**M**

114

.....

-----  
STEP SIX            EXIT HAP  
-----

The heading listing the modules has output on the terminal. The analyses are complete and you are ready to exit from the program by calling the 'STOP' module.

THE TERMINAL INPUTS AND OUTPUTS FOR STEP 6 ARE:

```
*****HYDROLOGIC ANALYSIS PROGRAM*****  
**ENTER: BOUNDARY, HYDROLOGIC MODEL, UPDATE, LOAD, STOP**  
[STOP]
```

STOP    NORMAL HAP EXIT

These interactive inputs may be placed in a data file prior to HAP execution. HAP can be instructed to receive inputs from this file creating a quasi-interactive terminal session. This mode is especially advantageous when testing the same watershed for many different rainfall rates.

#### CONCLUSION

Because of its interactive, real-time capabilities, the Hydrologic Analysis Program (HAP) is very valuable as a tool to assist in decision making processes. Although it is optimized for water resource analyses, HAP's ability to access any area or cell within the MSDAMP data base and then produce statistical tabulations or symbolic maps should make it an extremely attractive tool for an array of problems encountered by county officials. HAP is currently operational on the MNCPPC's HP3000 mini-computer and the UOMCE UNIVAC 1180 main frame computer.

#### REFERENCES

1. U.S. Army Corps of Engineers, 'Guide for the Creation of Grid Cell Data Banks', Hydrologic Engineering Center, USACE, Davis, Calif. Sept. 1978.
2. U.S. Department of Agriculture, Soil Conservation Service, SCS-TR-55
3. U.S. Department of Commerce, Technical paper 40

## AN INTERFACE FOR REMOTE SENSING DIGITAL IMAGE SYSTEMS AND GEOGRAPHIC INFORMATION SYSTEMS

R. R. Irish and W. L. Myers  
The Pennsylvania State University  
University Park, Pennsylvania 16801, U.S.A.

### 1. INTRODUCTION

Rapid developments in computer technology and applications programs have made possible the successful classification of a variety of regional natural resource phenomena by computer analysis of remotely sensed data. The same developments have engendered sophisticated polygon-based geographic information systems (GIS) for handling environmental and natural resources data. However, these two technologies have evolved separately, and an absence of interfacing has resulted. This is attributable, in large part, to different system concepts in representing space. Nonetheless, it is possible to establish a linkage between the two.

### 2. POLYGON AND RASTER DATA STRUCTURES

Polygon-based GIS's employ vector data structures in representing space. Areal entities are geocoded as an aggregation of polygons, where each polygon represents a homogeneous area of a map. An area's boundaries are commonly encoded as a circuit of X-Y coordinates. Encoded polygons are accompanied by unique reference codes that identify the polygon and serve as relational links to records in tabular files stored in the data base. These files consist of polygon descriptors--attributes that depict an aspect of the encoded map. The polygons and codes are stored in the data base as a file of polygons and, when linked to a set of attributes, constitute a layer of information.

Common borders of neighboring polygons are redigitized when polygons are encoded independently. Redundancy may be avoided by encoding boundaries delimited by nodes (points where three or more lines meet) and independent reference points, with one point in each polygon. The boundaries and points are correlated using a chaining algorithm. This results in lists of line segments composed of right and left polygon identifiers (two nodes and any number of points). Attribute values may replace identifiers for subsequent analysis and display.

Remote sensing scanners generate data in a raster format. The earth's radiant flux is recorded in two dimensions as sensing optics repeatedly scan the earth's surface in a sweeping motion perpendicular to the platform's orbit path. Telemetered data are put onto a computer-compatible tape in the format of a digital image data set. The raster-structured digital data are a matrix of spectral reflectance values, where each row represents a scan

line and each cell or pixel of the matrix is composed of a series of bytes, one for each wavelength as recorded by the scanner. The data set undergoes spectral pattern analysis, in which each pixel is assigned a symbol that identifies the earth surface category of which it is a constituent. This output is called a classified image file.

### 3. INFORMATION TRANSFER ALTERNATIVES

The desire to make use of pixel classifications in a GIS requires a means for making vector-based and raster data structures compatible. Either raster-to-vector or vector-to-raster conversion is necessary.

Raster-to-vector conversion results in a distinct polygonal layer for loading into the GIS data base. This may be accomplished by outlining feature category boundaries on a hard-copy pixel display and encoding the graphics. However, cleaning problems associated with table digitizers become part of the process, and automatic digitizers are expensive and scarce. The human becomes a decisive ingredient in an otherwise machine-oriented environment.

Raster-to-vector algorithms have been written for converting raster-based data to polygonal layers [Morehouse and Dutton 1980, Nichols 1982]. Implementation problems arise, however, due to substantial memory requirements placed on a computing system. Peripheral storage may also be limiting when deriving additional layers from successive data, and computing costs are high.

A second interfacing possibility involves relating classified pixels to an existing GIS polygonal layer. This is accomplished by rasterizing the layer into a classified image format. Once the rasterized version is created, it can be integrated with a classified image file by means of a digital overlay. This concept was used in developing a vector-raster interface that transfers classified image information to a GIS data base.

### 4. ZONAL INTERFACE

#### 4.1 Approach

The developed interface relies on existing polygonal layers from a GIS and is entitled Zonation Algorithms (ZONAL). Ownership parcels, political boundaries, administrative subdivisions, forest management compartments, and other geographic layers comprise the spatial data base of a GIS. These pre-defined polygon files serve as the geobase for numerically overlaying the classified pixels.

The digital overlay is accomplished by first rasterizing a file of polygons. A computer-simulated scanner generates a grid cell representation of a polygonal layer--a spatial replica of the Landsat information. However, the information content differs. Instead of a spectral reflectance value, each newly created pixel has affixed to it a code that identifies the polygon within which the center of the pixel falls. This results in an indexing scene linking each image pixel with a polygon from the GIS (Figure 1).

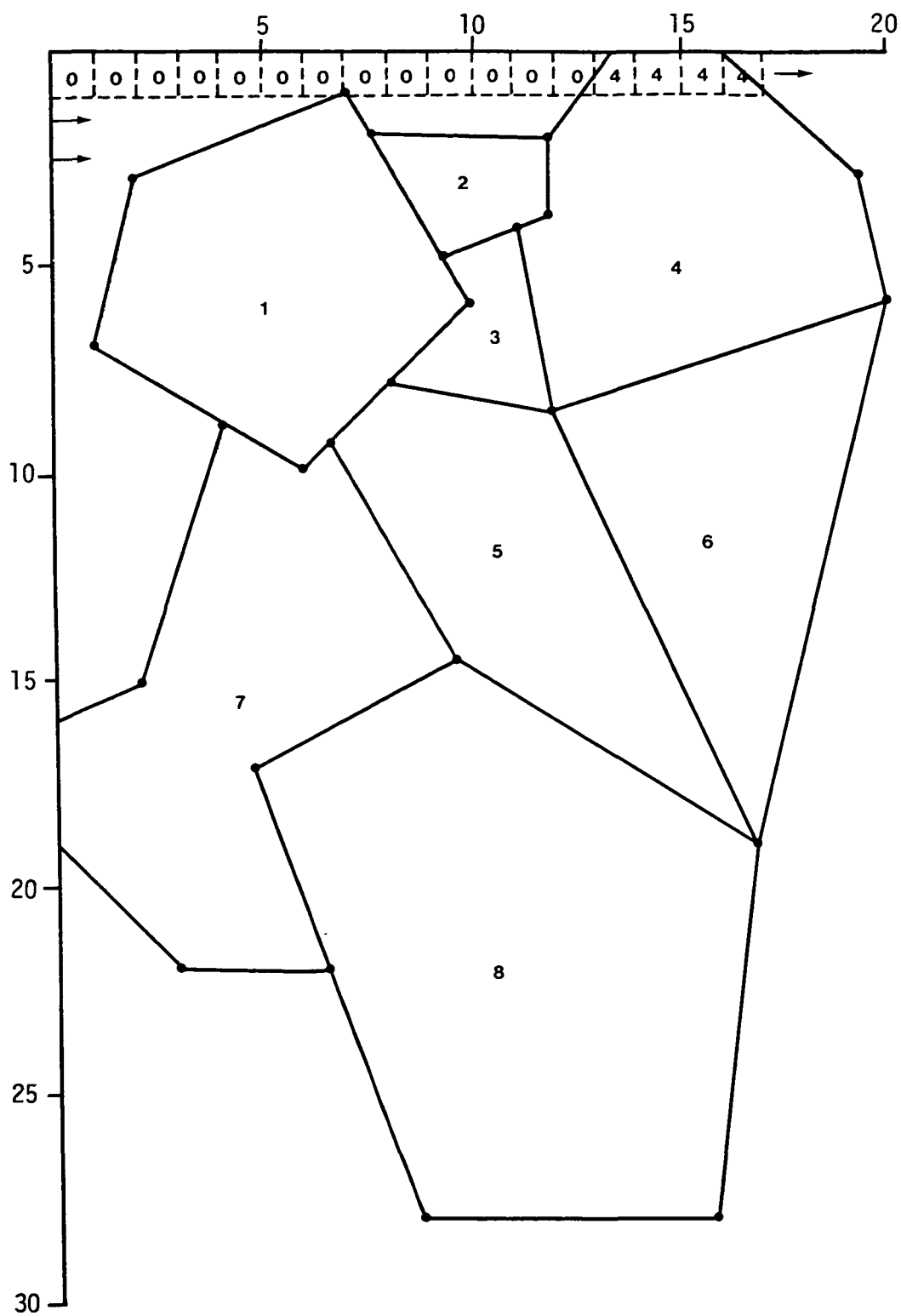


Figure 1. Illustration of how a polygonal layer is converted to an indexing scene. Polygon identifying codes are assigned to pixels on a scan line-by-scan line basis.

Classified pixels are tabulated within each polygon by simultaneously processing the classified image file with the indexing scene. The tabulation results take the form of frequency distributions depicting the number of classified pixels by feature category within each polygon. An analysis is performed on the frequency distributions, resulting in assignment of a feature category percentage figure or a label representing a co-occurrence of feature category percentages to each polygon. The resulting classification of polygons is formatted as either a numeric or non-numeric attribute file and, when linked to the polygonal layer, represents an additional layer of information. This file is transferred to the data base using the update facilities of the GIS.

#### 4.2 Host System Requirements

The ZONAL interface requires that GIS spatial layers be encoded as independent polygons or chains. Independent polygons must be stored or extracted as layers without internal overlap. Chains comprising a polygonal layer must be retrieved and stored as a separate file. Also, all layers must be error free.

Several requirements are also placed on the remote sensing digital image system. Pixels generated by the image classifiers must be held in peripheral storage, because this file is used in the digital overlay. Due to variations in remote sensor altitude, attitude, and velocity, digital image data are not in positional agreement with polygon files in a GIS. Geometric correction facilities are necessary to ensure reliable indexing.

#### 4.3 Interface Description

The ZONAL interface is composed of eight Fortran programs. Their relationships to the host systems are illustrated in Figure 2. A short description of each program follows:

POLSEG takes the polygon- or chain-based files as defined in a GIS and decomposes them into line segments.

ORDSEG orders the line segments into user-defined panels as a preprocessing step to rasterization. Ordering enhances the efficiency of pixel generation and reduces memory requirements for large polygon files.

GENPIX creates an indexing scene from the file of ordered line segments. Index pixels are generated on a panel-by-panel basis.

RIDPIX digitally overlays the classified image file with the indexing scene. RIDPIX determines the image data spatially coincident with the indexing scene and rewrites this registered subset, suppressing all non-polygon image information.

COMPIX overlays the registered image file with the indexing scene. Simultaneous processing results in frequency distributions depicting the number of classified pixels by feature category within each polygon. An output listing provides the polygon area covered by each classification category.

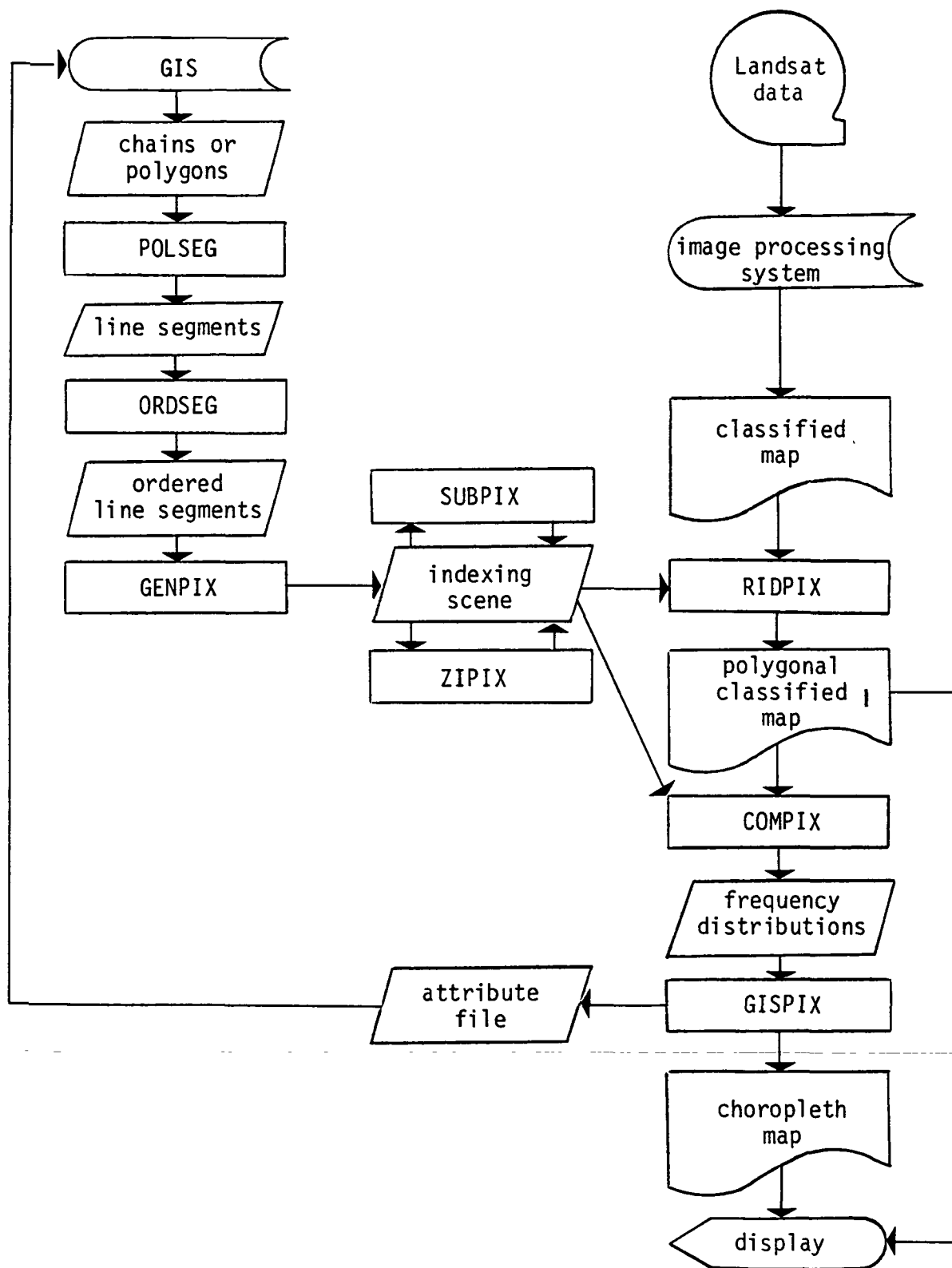


Figure 2. ZONAL flow chart.

GISPIX performs a secondary classification on the frequency distributions. Numeric or non-numeric attribute files are created by assigning a classification percentage or a label representing a co-occurrence of classification percentages to each polygon. An option generates a choropleth map.

SUBPIX extracts subareas from a larger indexing scene.

ZIPIX joins two indexing scenes along a common boundary to form one continuous indexing scene.

The ZONAL interface involves a multi-step operation. Six of the eight programs (POLSEG, ORDSEG, GENPIX, RIDPIX, COMPIX, and GISPIX) are necessary and must be used sequentially. SUBPIX and ZIPIX complement the interface by simplifying more complex indexing situations.

## 5. TEST CASE

To demonstrate the utility of ZONAL, a linkage was made between two host systems at The Pennsylvania State University. PSU's Experimental Forest, comprised of nine management blocks, was used as a study area.

A Task Oriented Multi-purpose Information System (TOMIS) is under development at PSU's School of Forest Resources [Myers 1982]. TOMIS, a polygon-based GIS, was designed to handle and analyze data associated with management- and research-related activities of experimental forests. Polygons are independently encoded as circuits of X-Y vertices. Attributes reside as either numeric or non-numeric descriptors and each is composed of two parts, an attribute type and value.

The digital image processing system developed by the Office for Remote Sensing of Earth Resources (ORSER) was used as the host image analyzer. A portion of a Landsat scene covering the Experimental Forest and scanned in May of 1973 was classified and geometrically corrected using the ORSER software [Turner et al. 1982]. Five land use and cover classes (water, coniferous forest, deciduous forest, senescent vegetation, and agricultural land) were defined. The senescent category comprises all areas of pre-leaf vegetation in the spring data set.

Nine polygons, representing the Experimental Forest's block boundaries, were assembled into and stored as a polygonal layer. The layer was decomposed by POLSEG into line segments. ORDSEG ordered the line segments within five panels. The indexing scene created by GENPIX involved two steps. First, the size of the raster file necessary to cover the management block layer at a given resolution (Landsat pixel) was determined. The panels were then processed sequentially, and each index pixel was assigned a code that identified the management block within which it fell.

At this point, the indexing scene and classified image file were processed simultaneously by RIDPIX. Based on a ground control point specified in terms of digitizer X-Y coordinates and image row and column positions, the subset of positionally coincident image pixels was determined from the parent file. The two were digitally overlaid, resulting in a rewritten, registered subset with



all image information outside the management blocks suppressed. This file was displayed and registration accuracy visually verified.

The registered file and indexing scene were digitally overlaid by COMPIX. Frequency distributions depicting the number of classified pixels by feature category within each management block and an acreage listing were generated. These were useful in evaluating and comparing the land use and cover classes as they occurred in the blocks. However, worthwhile feature category relationships existed among the pixel summarizations which were concealed in a tabular format. GISPIX made detection of these possible by examining the frequency distributions and extracting both numeric and non-numeric TOMIS attributes. Non-numeric attributes resulted from a second analysis of the image data in which management blocks were classified by recognizing co-occurrences of pixel percentages.

The simplest attributes consisted of percentage values of a single feature category. A request was made for coniferous cover attributes. The symbol representing conifers, the range of acceptable percentage limits (0-100 percent), and an attribute type (CONIFERS) were specified. The frequency distributions were processed and the attribute file created. Each record consisted of the block's code, the CONIFERS attribute type, and the percentage of the block covered by coniferous forest. A similar request was made for forested cover attributes. The two symbols representing the forest categories, the valid ranges for each category (0-100 percent), and an attribute type (FORESTED) were specified. In this case, each attribute record consisted of the block's code, the FORESTED attribute type, and the percentage of the block covered by forest.

A non-numerical set of attributes depicting the nature of each block's forest cover was derived by polygon classification. Criteria were established for assigning one of three attribute values to the management blocks: DECIDUOUS, CONIFEROUS, or MIXED forest. Polygons were classified on the basis of the following set of criteria:

Range Limits (%)		Threshold Percentage	Attribute Type	Attribute Value
Deciduous	Coniferous			
0 - 9.9	65.1-100	75	FOREST	CONIFEROUS
65.1-100	0 - 9.9	75	FOREST	DECIDUOUS
10 -100	10 -100	75	FOREST	MIXED

Any blocks with less than 75 percent forest cover were not assigned a value. All blocks containing over 65 percent coniferous or deciduous forest were assigned the appropriate value, provided the remaining forest cover was less than 10 percent and the 75 percent threshold was met. All other blocks with over 10 percent coniferous forest were assigned the MIXED forest value if the total forest cover was at least 75 percent. The decision regions and results of classification are portrayed in Figure 3.

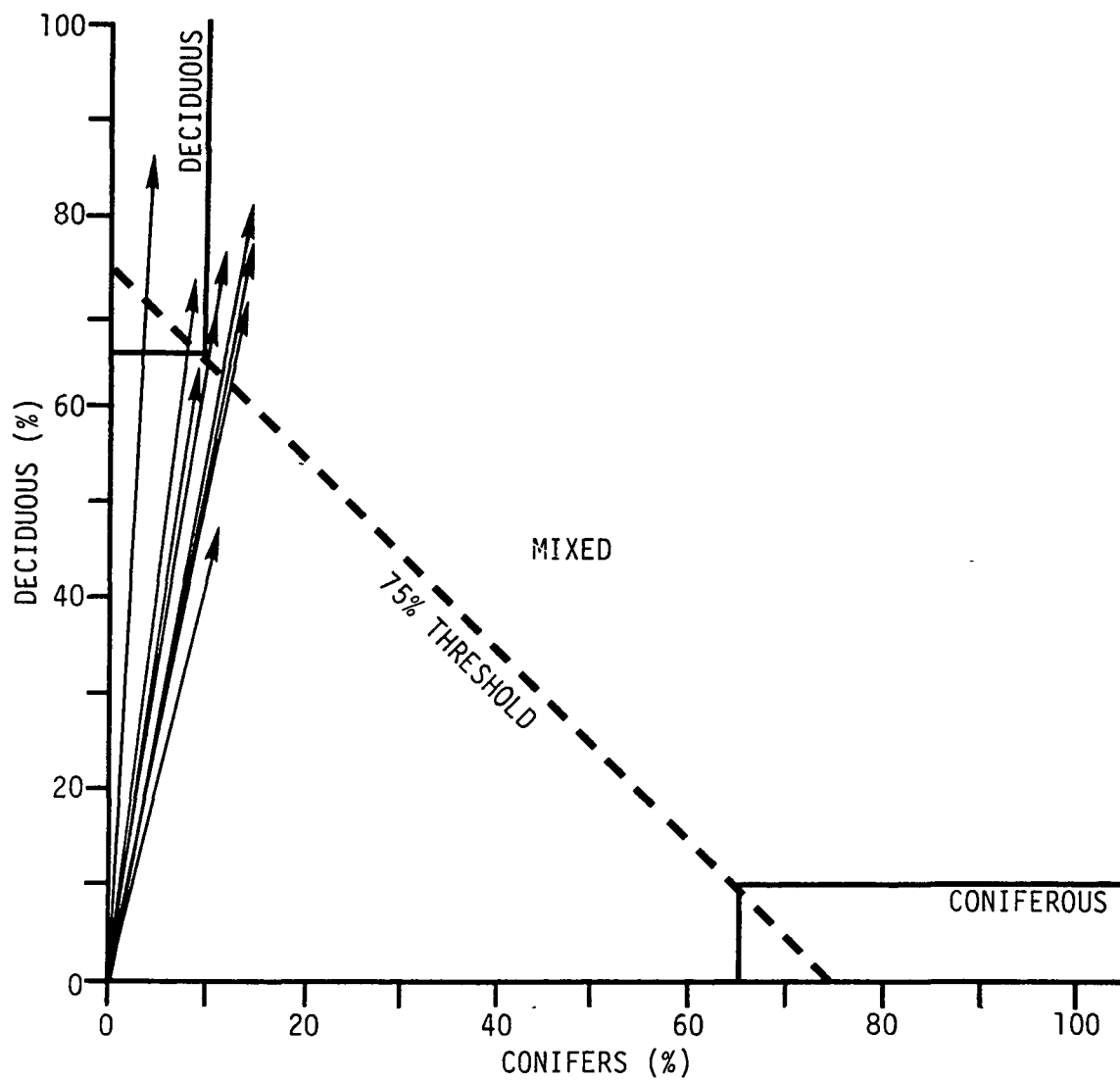


Figure 3. Decision regions and results of polygon classification. Vector endpoints portray how each management block was classified.

The preceding application serves to illustrate the ZONAL interface; however, indexing Landsat information to limited land areas is probably impractical. A unique aspect of analyzed remotely sensed data is the global perspective provided in recognizing occurrences and distributions of earth surface phenomena. Also, GIS's are typically employed to store and analyze detailed polygonal layers covering extensive land areas. More realistic applications would include inventorying and monitoring forest resources over industry-owned lands, surface mine evaluation and change detection analysis, and county-based land use and land cover inventories.

## 6. CONCLUSIONS

The ZONAL interface offers a reasonable means of utilizing Landsat information in a polygon-based GIS. The ZONAL mechanisms for information transfer are based on the use of existing GIS polygonal layers, thereby making the process entirely automated. The indexing scene permits location-specific inventories of terrain features within selected polygons. Through indexing, a voluminous set of image pixels is condensed to frequency distributions. By polygon classification, numbers buried in summarization tables can be extracted and analyzed. In this way, relationships are identified and polygons meaningfully characterized. GIS storage requirements are lessened by entering summations of relevant classification categories. Once an indexing scene is created, it may be used repeatedly in keeping a data base current, provided the polygonal layer remains unchanged. Additionally, ZONAL can be adapted to other processing systems and GIS's because host system modifications are not necessary.

## 7. LITERATURE CITED

- Morehouse, S. and G. Dutton. 1980. Extraction of polygonal information from gridded data. Pages 320-327 in R. T. Aangeenbrug, Ed., AUTO-CARTO IV, Vol. 2. American Congress on Surveying and Mapping and the American Society of Photogrammetry, Falls Church, Va.
- Myers, W. L. 1982. TOMIS: Task Oriented Information System Users Manual. School of Forest Resources, The Pennsylvania State Univ., University Park, Pa. (Mimeographed.)
- Nichols, D. A. 1982. Conversion of raster coded images to polygonal data structures. Pages 508-515 in B. F. Richardson, Ed., Proceedings of the Pecora VII Symposium. American Society of Photogrammetry, Falls Church, Va.
- Turner, B. J., G. M. Baumer, and W. L. Myers. 1982. The ORSER Remote Sensing Analysis System: A User's Manual. Res. Publ. 109/OR, Institute for Research on Land and Water Resources, The Pennsylvania State Univ., University Park, Pa.

## DEVELOPMENT OF LANDSAT DERIVED FOREST COVER INFORMATION FOR INTEGRATION INTO ADIRONDACK PARK GIS

R. Curran and J. Banta  
Adirondack Park Agency  
Ray Brook, New York; U.S.A.

Based upon observed changes in timber harvest practices partially attributable to forest biomass removable for energy supply purposes, the Adirondack Park Agency began in 1979 a multi-year project to implement a digital Geographic Information System (GIS). An initial developmental task was an inventory of forest cover information and analysis of forest resource change and availability. While developing the GIS, the Agency, with consultant assistance, undertook a pilot project to evaluate the usefulness of Landsat derived land cover information for this purpose, and to explore the integration of Landsat data into the GIS.

The prototype Landsat analysis project involved 1) the use of both recent and historic data to derive land cover information for two dates; and 2) comparison of land cover over time to determine quantitative and geographic changes. The "recent data," 1978 full foliage data over portions of four Landsat scenes, was classified, using ground truth derived training samples in various forested and non-forested categories. This inventory resulted in the classification of 83 percent forested and 17 percent non-forested, as generalized by combining categories. Forested categories include the following: northern hardwoods, pine, spruce-fir, and pine plantation, while non-forested categories include wet-conifer, pasture, grassland, urban, exposed soil, agriculture, and water.

Similarly classified "baseline" data (1972 leaf-off data) were found to be generally incompatible with the "recent" classification because of an overestimate of non-forested areas. A conservative interim estimate of forest cover loss over the period was, however, derived by narrowing the evaluation to one classifier which detected in the "recent" data forest areas subject to cover loss due to harvest, pests or blowdown. Areas classified by this signature were compared to the baseline data, to determine whether or not they were forested in 1972, using a 9 pixel search window.

Lower quality, but full foliage, baseline data is currently being processed to provide a more comprehensive analysis of forest cover infill as well as loss.

The 1978 landcover information has been integrated into the Agency's GIS as one of three study area complete data sets. Digital geographic data is stored in raster format on a one acre grid cell structure keyed to the New York State Universal Transverse Mercator grid. Data is digitized and processed using a turn key packaged micro-processor system purchased by the Agency. Using the GIS, land cover data for selected municipalities has been combined with other variables and used to measure development proximity to "critical natural sites", predict recreational use potential, and estimate the regulatory protection afforded forest and open space resources by Agency planning. New digital data sets for soils, elevation, public infrastructure, economic and demographic data are planned for 1982-84 addition to the GIS. These will expand the potential for forest cover analysis by adding site considerations, either by providing a weighted variable in the Landsat image processing stage of analysis, or in geographic comparisons of raster data sets using other GIS capabilities.

## INTRODUCTION

### Agency Role in Adirondack Park

The NYS Adirondack Park is an area of approximately 9300 square miles in northern New York. It is composed of intermingled large state and private landholdings with scattered small settlements ranging in size to about 7,000 residents. About one-third of the land is state-owned. Virtually all state-owned land is removed by state constitutional rule from timber harvesting.

Large private landholdings are held by relatively few landowners. About 50% of the private land is in holdings greater than 1000 acres in area. The bulk of this land is devoted to forest management by industrial landowners. Some is held in private preserve status which may also be managed for timber and biomass production.

The Adirondack Park Agency is an Agency of the NYS Executive Department with regional responsibilities for planning and regulation of new private land-use and development. A sister agency, the NYS Department of Environmental Conservation, is responsible for the care and custody of most state-owned lands and state-wide environmental regulations. The New York State Energy Office supervises preparation and implementation of the "New York State Energy Master Plan II," now in its second interaction. The Park Agency works closely with other state agencies for differing objectives such as the development of the State Energy Master Plan, the policy document that guides all state agencies with respect to state energy policy (1).

### Reasons for interest in landcover information

The Adirondack Park Agency has sought techniques to rapidly assess natural resources park-wide and to track stress and significant changes over time. Primary short-term concerns of the Agency relate to responsibilities for permit issuance for new land use and development, and therefore to human or development-caused changes. Most clearcutting and shoreline cutting requires permits from the Agency. Other timber harvesting does not. Longer range Park policy concerns include forest diseases and acid precipitation, albeit in conjunction with other state agencies that have principal responsibility for such issues state-wide, especially the Department of Environmental Conservation.

Both the managed forests and the Forest Preserve (the reserved state lands) are subject to a variety of forest disease problems such as scleroderris in red pine, beech bark disease, and others. These problems diminish the attractiveness of infested stands for traditional markets, and have further stimulated interest in biomass market for energy. With a market for low-grade material timber stand improvement could become more economically attractive (2).

The assessment of public and private forests for real property taxation has become an issue as fiscal constraint and court-mandated changes in assessment procedures are reflected in the property tax system. Upward pressure on forest land real tax burden is reflected by tax increases in some towns in the Park on the order of 100-150 percent (3).

Private forests, at least prior to the current economic recession, have also been subject to increasing cutting pressure for home fuelwood, and conventional pulp wood production (4).

Interest in Adirondack biomass resources reached one peak with the proposal by the City of Burlington, Vermont to rely on wood-fueled electric power generation for a significant portion of its electric energy supply. The Burlington proposal is based in part on a smaller-scale experimental program that utilized as one source of fuel green-wood chips produced for the city as a by-product of an operation feeding wood-chip biomass from the Northern Adirondacks to a paper mill in Cornwall, Ontario (5).

The Burlington proposal has since been changed to rely heavily on rail transport of chips, a factor which will significantly limit the supply that could be shipped from the Adirondack Park given its existing rail network.

The State Energy Plan documentation also reveals a number of northern Adirondack biomass energy proposals, some of which go beyond the experimental stage. Clarkson College in Potsdam, New York has adapted a major boiler facility for green-wood chips (6).

One proposal for an experimental wood-fired electricity and steam generation plant in Tupper Lake, New York failed to make the grade with grant reviewers, apparently because of a lack of market for cogenerated steam and uncertainty about the availability of local low-grade biomass (7).

Thus, the shift to biomass energy within the Adirondack Park appears limited to its use as home fuelwood for the near future. A change in general economic conditions could, however, unleash general demand for the wood biomass resources of the region, leading to consequences that are significant for the Adirondack Park policies guided by the Adirondack Park Agency. Under assumptions of good forest management, this could lead to general improvements in productivity and profitability of a major industry in the region over the long term. Less optimistic seers worry about large-scale clearcutting and associated problems for regeneration of valuable forest stands, site specific environmental disruption and dependence on chemical treatments to deal with nutrient losses or protection of planted species from pests or competing vegetation.

The Agency examined these questions at some length in 1980 and 1981, resulting in several special reports summarized in a

publication "Clearcutting in the Adirondack Park", a report of the Joint Government-Industry Steering Committee on Intensive Timber Harvesting in the Adirondack Park to the Adirondack Park Agency, May, 1981 (8).

The Agency's policy concerns were complicated by lack of year-by-year data that correlated well with the 1968 or 1980 USFS surveys of standing timber supply. Accurate projections of growth and removals as they might be influenced by biomass removals have not progressed beyond forecasts of available supply (9). Stocking information is dated or lacking for the region except for some areas of industrial holdings where the information is considered proprietary.

The Agency commissioned a preliminary analysis using computer projection techniques from 1968 data along with phone survey and NYS Department of Environmental Conservation product removal information (10). This showed significant drains on hardwoods, but has proved difficult to reconcile with 1980 USFS survey data because of uncertainty about fuelwood biomass harvest and inherent difficulties in projecting forward from the USFS inventory data. Preliminary 1980 survey statistics suggest that the computer projection under estimated available inventory (11).

At the same time in 1981 the Agency made final commitments to a computerized Geographic Information System (GIS). The GIS as a first priority is intended to extend and document McHargian overlay analysis of development constraints within the Park adding economic and demographic factors to allow a fuller documentation of the Adirondack Park Land Use and Development Plan, a regulatory plan aimed at new land use and development of regional significance within the Park (12).

The GIS system for the Park was assembled in a configuration that would also permit direct Landsat image processing, both with the NASA Landsat satellites and the upcoming Landsat D technology. While this is of interest to the McHargian development capability analysis, it is of primary interest because of the ability to track larger-scale resource changes on a park-wide basis.

The concern for biomass removals and the Agency's legal requirement for a permit for clearcutting as defined in the Adirondack Park Agency Act, caused the first applications of the Landsat change analysis techniques to be made to determine forest cover change. Similar legal responsibilities address wetlands, and the dynamics of the wetlands systems of the Park are of equal ecological significance, but have been deferred to the completion of a remapping of the wetlands in the Park to 1 acre size thresholds (state-wide mapping uses approximately 12 acre thresholds) (13).



The prototype Landsat analyses addressing biomass availability and clearcutting are described in the following sections of this paper.

#### LANDSAT LANDCOVER ANALYSIS

The Agency's first goals with regard to the use of Landsat data were to investigate the usefulness of these data in the varied land use planning analyses needed by the Agency. Reliability, cost effectiveness, timeliness of coverage, labor required and ease of access are limitations associated with conventional photo interpretation. Landsat promised to reduce the cost of rectification of these liabilities through digital analysis.

To address the issues of biomass availability and forest cover change, the prototype Landsat analysis would seek to produce a classification of recent data, including classification of signatures for cut forest land, for areas of new construction or disturbed soil, for forested land, wetlands and developed or open areas. In addition to examining the potential for determining areas with changing landcover by their characteristic signature, a temporal change analysis was to be undertaken to compare current landcover data to past conditions. The temporal analysis would permit an estimation of forest infill as well as loss. Underlying these expectations was the need to produce a relatively simple classification, in a cost effective manner, which could be duplicated at the end of another period in order to continue to monitor landcover trends. Cartographic geographical registry with other GIS data bases is a secondary issue of particular significance to these applications.

#### Methodology

A supervised classification using training samples typical of the cover type to be classified was undertaken. Because of its size the approximately 6 million acre Adirondack Park offers a spectrum of landcover sites from which to choose training samples; however, the study area is split by four Landsat scenes (Path 15; Row 29, 30 and Path 16; Row 29, 30). An examination of available data revealed that during the period 1970 to 1979 very few acceptable data choices were open mainly because of cloudy conditions in scenes recorded. The dates chosen for a recent data set (Path 15 - August 22, 1978 (Scene I.D.'s 83017015011, 83017015014) and Path 16 - June 30, 1978 (Scene I.D.'s 83011715061, 83011715064) enabled cloud free full foliage data within a single season. However, the only cloud free data available at the beginning of the period was leaf off data for October 10-11, 1972. Initially, use of cloud free data was regarded as a priority concern and an attempt made to compare results using full foliage versus leaf off data.

The proposed digital analysis (a supervised classification) was to be undertaken by a contractor with the Agency determining training samples and verifying the classification results.

### Training Samples

Training samples were determined in each Landsat scene within the study area. These included forested type such as hardwoods (beech-birch-maple and birch-aspen types), spruce-fir, pine (as well as pine plantations), wet conifers (black and red spruce, larch and white cedar lowlands) and cut hardwoods (crown cover reduced by approximately 30%). No attempt was made to separate mixed woods. Non-forested training samples included areas of water, deep soil excavation, residential and commercial development, wetlands, grassland, brushland, pasture and agricultural cultivation.

Training sites were determined by using field verification of 1978 1:24,000 scale panchromatic photography to select areas representative of the cover type. Where possible the same sites were used for training on the 1972 images if photo verification revealed that no cover change had occurred on the site during the period covered by available photography (1968 and 1978).

### Evaluation

Completion of the final scene classification was an iterative process, involving several digital classification runs on each scene followed by field/photo evaluation and entry of new training sites, until a final product representing an acceptable interpretation of the data was derived.

After each supervised classification was produced, the areas covering 4 to 5 7½ minute quadrangles were extracted from the four classified images. Quadrangles were systematically checked for accuracy to determine consistent errors and to identify areas for new training sites.

Ground truth verification revealed the following characteristics of the eventual 1978 classification.

- 1) Areas classified as conifer were conservative.
- 2) The hardwood classifier included mixed hardwood-conifer situations.
- 3) Emergent wetlands could not be differentiated from mesic grass lands (apparently because of dry conditions).
- 4) The pasture classifier included occasional houses and lawns in woodlands, hardwood woodlands disturbed by cutting, blowdown, or pests.

5) Wet shadowed bedrock out croppings on steep slopes were classified as water.

6) Pine consistently identified both pine and hemlock while spruce-fir included pine as well as spruces and fir.

7) The pasture, brush, exposed soil, agricultural and urban categories consistently occurred in areas with a disturbed landscape.

### Statistics

The 1978 classification revealed that 83% of the Adirondack Park was forested with 17% classified as non-forested. Forested categories included northern hardwoods (56%), spruce-fir (19%), pine and pine plantation (8%).

Because the classification of the October, 1972, data (which eventually involved the use of approximately 50 training sites and 3 classification iterations) consistently underestimated forested areas, particularly hardwoods, use of that data for a complete landcover classification was abandoned. New full foliage data was secured for other dates and is currently being processed.

Useful comparisons could, however, be made between the 1972 and 1978 data by limiting the categories compared. Changes from forested in 1972 (a conservative estimate of forested areas) to pasture in 1978 (the classifier which included disturbed forest land) were found to accurately delineate forest land disturbance due primarily to timber cutting. A 9-pixel moving window comparison was used to determine these changes in landcover. Field verification showed that this technique identified areas of clearcutting for timber removal and clusters of new buildings.

### Geographic Information System

As the landsat landcover project demonstrated its usefulness by providing rapidly available and reliable data to the Agency; a GIS, in part, to accept and enhance the Landsat data was designed and implemented.

Five primary information management goals were identified for the digital component of the GIS including: 1) digital data storage (in raster format); 2) planning analysis capabilities; 3) the ability to rescale mapped data; 4) access to remote data sources and 5) Landsat processing capabilities. The computer operated component of the GIS was designed to enhance existing manually stored information and to prepare hardcopy maps which were compatible with the existing map file. The system was to be constructed under contract and delivered in working order with certain basic data entered to the Agency.

## Hardware

The system delivered consists of 1) a micro-computer with 64kB of memory as the central processing unit of the system; 2) a custom designed hard disk system having a total of 96 megabytes of storage, of which 80 mb are contained in a non-removable disk and 16 mb in a removable disk pack; 3) a 1600 BPI tape drive unit which serves as a data backup and rescue to the disk drives (data can be moved between the system and remote sources using the tape drive unit or floppy disks); 4) a dot-matrix printer with a wide range of symbols and characters which provides grey-scale mapping and will print a 7½ minute quadrangle map at scale in 2 sections; 5) a color monitor with 128 kB of memory, capable of displaying 512 by 480 pixel image.

Mapped data can be entered from a variety of sources using a 54" x 36" digitizing tablet.

## Software

Because the GIS is operated by and serves a diverse professional staff and lay person constituency untrained in programming skills the software operating the GIS is user-friendly and operates in an interactive mode. The software does however offer the opportunity to communicate with other remote digital data bases which do not necessarily operate in a user friendly atmosphere.

A raster system for software operation, as opposed to a polygon, was chosen because of its lower cost for data storage, equipment acquisition, and analysis. The software performs planning analyses including multi-variable geographic combination, coincidence of variables, capability analysis and weighted variability analysis. The choice of raster format effectively limits the usefulness of the system to produce certain types of line oriented cartographic products from stored data.

Data can be accessed on the basis of a micro grid structure based upon the 7½ minute series quadrangle maps which coincides with the Agency's manual map file system. Analysis of data, including Landsat processing, is most effectively accomplished on a small area basis (e.g. quad by quad); analytic results can then be stored within the macro-grid the overall Park area, for further aggregation and analysis.

Additionally the software is designed to incorporate a data base management methodology which will enable access to complex data variables and an attribute file geographically keyed to the content of the raster data file.

## Data Entry and Storage

The GIS raster system is based upon a square 1 acre cell structure. This size allows an accurate portrayal of the Park Plan Map, the land use area map forming the basis of the regulatory and planning program of the Agency, and is compatible with Landsat data. By using the NYS UTM grid to define the perimeter of each cell, the grid cell structure established in the GIS can be easily referenced to the Agency's existing cartographic data base, especially the USGS 7½ minute quadrangle series, for data entry and retrieval.

Three park-wide data variables, including private and state land classifications, unique natural and cultural features, and land cover data from 1978 Landsat overflights, comprise the initial system data. A fourth parkwide data base, political boundaries stored in polygon four, is used in conjunction with other data.

Data in each variable entered is in a 4 bit format, with the option of using an 8 bit format for a variable file which exceeds 15 subclasses. File headings describe the content and location of each variable file and identify associated attribute files developed to enhance the information available for description of variables at a given site. Each Parkwide gridded in 4 bit code occupies 7.5 mb 7 disk storage space. Using a 4 bit code, the 80 mb fixed disk pack system will have the capacity to store 9 or more parkwide data files, on line at one time. Additional data files, two per removable disk pack, can be accessed on the removable disc pack channels for increased flexibility.

These data files are central to Agency concerns about forest cover change because they relate the administrative and regulatory rules of the Agency to observed changes in land cover in the land cover files.

## Land Classification and Boundaries

The classification of land on the Adirondack Park Land Use and Development Plan Map (APLUDP Map) and the State Land Master Plan Map (SLMP Map) were digitized from 199, 7½ minute series quadrangle maps contained in the Agency's manual file system. These maps include UTM reference coordinates and contain 15 sub-classes.

The Adirondack Park Land Use and Development Plan Map, one of two major elements, incorporated into the "Classification of Land" GIS file, describes the classification of all private land in the Park into discrete land use areas. Three major land capability factors determine these land use areas including the natural resource amenability to development, the level of available public services and the open space attributes of the land in question

(14). There are seven (7) classifications on the Park Plan Map, such as Hamlet, Industrial Use, and Resource Management.

In addition, the SLMP Map shows nine (9) categories, such as Intensive Use and Wilderness, for all state lands of the Park; these classifications define recreational use potential based upon existing uses, wilderness character and the capacity of the land to sustain various levels of recreational use (15).

Political boundaries entered as UTM vertices from a 1:250,000 series planimetric map include all minor and major municipalities within the Park, 112 counties, towns and villages. These data were converted to and are stored as study area cell vertices to be used to extract data from other Parkwide gridded data sets.

The Land Classification and Boundary data files will serve to define areas within which to aggregate and report data. Additionally, the classification of land has regulatory implications as well as associated policy goals applicable to the development or use of land within the Park. The classification of private land determines such regulatory requirements as the permitted density at buildout, the required minimum shoreline setback for new structures or the types of new development which would require "building permits."

A primary use of the data in relation to landcover can be to determine, using coincidence of variables operations the distribution of land classification entities (as acres) by Landcover type. This information would reflect, in a general fashion, the impact of Agency regulations on clearcutting as well as on other potential development.

Furthermore, as classifications are modified by amendment these aerial figures can be easily updated.

#### Unique, Natural and Cultural Features

The presence of Unique, Natural and Cultural Features were digitized as point data. An associated attribute file has been created from data currently stored in notebooks. Sites in the data file consist of key plant and animal habitats, rare and endangered species locations, key geologic sites, historic sites, and potential or existing hydropower sites.

Initially the attribute file information will be digitally analyzed to determine the distribution of sites by categories, such as rare species or waterfalls. Sites which are sensitive to disturbance will be identified and mapped for use in the Agency's regulatory program and the administration of state land units. Using a proximity operation sites in the vicinity of disturbed landcover can be identified as a measure of the environmental impact impinging on the sites.

## APPLICATIONS OF FOREST COVER DATA IN GIS

Because the landcover change analysis is incomplete, illustrations of GIS integration of forest cover change analysis relevant to biomass energy management have yet to be performed. At some point, a Parkwide data file containing areas of forest cover loss and infill will be created and utilized in multi-variable combinations in the GIS. Pending the availability of this change data, other applications of current data variables which are illustrative of the same combinations possible with forest change data have been developed. These comparisons, discussed below, utilize 1978 forest cover data in combination with other GIS data files.

The impact of varying degrees of land use regulations can be determined by an analysis of the coincidence of forested landcover types with categories of land classification. For example, the percentage of forested land which could be removed from production by urban uses can be estimated by determining the acreage of forested land in the land use categories of Hamlet, Moderate Intensity and Low Intensity, those areas with high densities of permitted development. Similarly, the same analysis can be performed using other proposed variables or existing GIS variables, such as Unique, Natural and Cultural Features sites. A prototype analysis performed for a Park County (Table 1) compares the landcover characteristics within approximately  $\frac{1}{2}$  mile of two categories of UNCF sites, to which differing regulatory protection goals relative to protection from a landcover removal would be applied. These are, first, historic sites (mainly reflecting buildings or registered sites) and secondly, natural areas (including habitats of protected plants or animals). Conversely this analysis demonstrates the comparative vulnerability of these categories to unregulated removal of landcover. A major limitation of the existing source data, that only the occurrence of a site in the location has been recorded rather than the extent of the feature, can be corrected by digitizing polygons of the features to delineate their area. This method was tested in a local application of the GIS data. The print data does provide an indication that a feature exists in the area.

### Locality Application

One of the 112 municipalities within the Park has cooperated in more detailed analytic work as part of its local planning effort. More detailed and varied data was digitized and entered into a gridded data base established for the Town of Duane, in Franklin County. These data elements included meso level soil boundaries, slope categories, UNCF polygons and travel corridors.

An interesting application was to determine the landcover characteristics of deerwintering areas as reflected in state resource maps (Table 2). Deer wintering areas are thought of mainly as sheltered coniferous areas; the analysis showed,

however, that extensive hardwood areas were included within the wintering areas. Assuming the initial boundary is accurate, the hardwood areas may be important for feeding or movement from one conifer patch to another.

Another application, developed by combining limitations of soil and slope categories with landcover, examines the availability of timber for harvest (Table 3). By adding other variables, such as road location, scenic and recreational values, and ownership patterns, the regulatory implications of varying proposed local regulations controlling logging on timber availability can be explored.

### Conclusion

Introduction of a GIS to a professional staff without computer or landsat processing experience points to the following conclusions:

- o timely access to both current and historic raw landsat data and its cost is a very significant consideration in its useful application.
- o user-friendly micro-computer technology is an effective means of extending staff effectiveness by automating labor intensive cartographic analysis if digital data such as landsat are available.
- o quantitative forest cover change analysis can be determined with a high degree of reliability in the northern hardwoods of the Adirondacks if high quality full foliage data is available. (Another NYS Agency is presently evaluating qualitative analysis using supplementary resource and economic data such as found in the Agency GIS and other state record systems).

The chief advantages of the GIS are the ability to generate inexpensive resource inventories for large land areas with a relatively high degree of reliability and accuracy. The digitized data is easily manipulated and compared to other data in the GIS. Its use as a tool to monitor landcover changes has not been fully demonstrated although a limited application realized some return.

Beyond the completion of the temporal landcover analyses, anticipated capabilities include the integration of new data such as soils and elevation into the GIS to help to increase the accuracy of the interpretation of landsat landcover data and provide additional variables in analysis. In addition the GIS data, particularly soil characteristics, could be used as a weighted variable in a Landsat classification run to increase the accuracy of classification.



however, that extensive hardwood areas were included within the wintering areas. Assuming the initial boundary is accurate, the hardwood areas may be important for feeding or movement from one conifer patch to another.

Another application, developed by combining limitations of soil and slope categories with landcover, examines the availability of timber for harvest (Table 3). By adding other variables, such as road location, scenic and recreational values, and ownership patterns, the regulatory implications of varying proposed local regulations controlling logging on timber availability can be explored.

### Conclusion

Introduction of a GIS to a professional staff without computer or landsat processing experience points to the following conclusions:

- o timely access to both current and historic raw landsat data and its cost is a very significant consideration in its useful application.
- o user-friendly micro-computer technology is an effective means of extending staff effectiveness by automating labor intensive cartographic analysis if digital data such as landsat are available.
- o quantitative forest cover change analysis can be determined with a high degree of reliability in the northern hardwoods of the Adirondacks if high quality full foliage data is available. (Another NYS Agency is presently evaluating qualitative analysis using supplementary resource and economic data such as found in the Agency GIS and other state record systems).

The chief advantages of the GIS are the ability to generate inexpensive resource inventories for large land areas with a relatively high degree of reliability and accuracy. The digitized data is easily manipulated and compared to other data in the GIS. Its use as a tool to monitor landcover changes has not been fully demonstrated although a limited application realized some return.

Beyond the completion of the temporal landcover analyses, anticipated capabilities include the integration of new data such as soils and elevation into the GIS to help to increase the accuracy of the interpretation of landsat landcover data and provide additional variables in analysis. In addition the GIS data, particularly soil characteristics, could be used as a weighted variable in a Landsat classification run to increase the accuracy of classification.

## REFERENCES

The Geographic Information System described in this paper is an ERDAS 400, a turn key system provided by ERDAS, Inc., Atlanta, GA. Funding for the system was provided in part by the National Park Service, U.S. Land and Water Conservation Fund.

- (1) Energy Plan
- (2)
- (3) Equalization and Assessment studies
- (4) DEC Firewood Survey
- (5) Policy Forum
- (6) Policy Forum
- (7) T.L. Plant Report
- (8) Firewood Survey
- (9) S-C Report
- (10) ESF Biomass Predictions
- (11) Binckley
- (12) APA Act
- (13)
- (14)
- (15)

### References

Adirondack Park Agency, 1973. Adirondack Park Land Use and Development Plan, Adirondack Park Agency, Ray Brook, N.Y. 31p.

Adirondack Park Agency, 1972. Adirondack Park State Land Master Plan, Adirondack Park Agency, Ray Brook, N.Y. 40p.

Table 1

DISTRIBUTION OF LANDCOVER WITHIN 14 PIXELS ( $\frac{1}{2}$  mile) OF  
UNCF SITES IN CLINTON COUNTY

Landcover Type	Historic Sites (% of area)	Natural Sites (% of area)
Hardwood	12	30
Spruce/fir	25	31
Pine	5	11
Pine Plantation	6	5
Wet Conifer	3	5
Pasture	11	3
Grassland	2	1
Brush	0	1
Agriculture	10	1
Exposed Soil	2	0
Urban	7	1
Water	17	13
Total Forested	51	82
Total Disturbed	32	5

Table 2

LANDCOVER FOR DEERWINTERING AREAS IN THE TOWN OF DUANE

Landcover Type	Acres	Percent
Hardwoods	5415.	46.06%
Spruce/Fir	3141.	26.72%
Pine	931.	7.92%
Wet Conifer	607.	5.16%
Pasture	235.	2.00%
Brush	32.	.27%
Grassland	7.	.06%
Agriculture	16.	.14%
Exposed Soil	0.	0.00%
Urban	13.	.11%
Pine Plantation	896.	7.62%
Water	463.	3.94%
Totals:	11756.	100.00%

Table 3

COINCIDENCE OF LANDCOVER/DEVELOPMENT LIMITATIONS  
IN THE TOWN OF DUANE

<u>Landcover Type</u>	<u>Degree of Physical Limitation</u>			<u>Total</u>
	<u>Moderate</u>	<u>Severe</u>	<u>Overriding</u>	
	(acres)	(acres)	(acres)	(acres)
Hardwoods	2,643	17,621	10,999	31,263
Spruce/Fir	1,450	6,930	11,531	19,911
Pine	343	1,809	703	2,855
Wet Conifer	177	1,123	409	1,709
Pasture	196	575	92	863
Brush	9	92	98	199
Grassland	31	22	32	85
Agriculture	68	80	4	152
Exposed Soil	4	0	4	8
Urban	38	31	12	81
Pine Plantation	403	1,907	110	2,420
Water	86	422	1,644	2,152
Totals:	5,448	30,612	25,638	61,698

## REFERENCES

The Geographic Information System described in this paper is an ERDAS 400, a turn key system provided by ERDAS, Inc., Atlanta, GA. Funding for the system was provided in part by the National Park Service, U.S. Land and Water Conservation Fund.

- (1) State Energy Office, 1981. New York State Energy Master Plan II and Long Range Electric and Gas Report, NYS Energy Office, Albany, N.Y. (3 Vol).
- (2) \_\_\_\_\_, 1980. Report of the Economics Subcommittee Adirondack Intensive Timber Harvesting, NYS Adirondack Park Agency, Ray Brook, N.Y., 143p.
- (3) Division of Equalization and Assessment, 1979. 1978 Revaluations Intra-Class Tax Shift Analysis (Part 1), NYS Division of Equalization and Assessment, Albany, N.Y., 11p.
- (4) Department of Environmental Conservation, 1981. The Forest Resources of New York, A Summary Assessment, NYS Department of Environmental Conservation, Albany, N.Y., 140p.
- (5) See Adirondack Park Agency, 1980. Intensive Timber Harvest Study Steering Committee Policy Forum, NYS Adirondack Park Agency, Ray Brook, N.Y. (2 Vol).
- (6) Ibid.
- (7) See Technical Assistance Center, 1977. Wood for Fuel Generating Electricity in Tupper Lake, N.Y., S.U.N.Y. at Plattsburgh, N.Y., 52p.
- (8) \_\_\_\_\_, 1981. Clearcutting in the Adirondack Park, NYS Adirondack Park Agency, Ray Brook, N.Y., 37p.
- (9) Department of Environmental Conservation, 1981. The Forest Resources of New York, A Summary Assessment, NYS Department of Environmental Conservation, Albany, N.Y., 140p.
- (10) \_\_\_\_\_, 1980. Report of the Economics Subcommittee Adirondack Intensive Timber Harvesting, NYS Adirondack Park Agency, Ray Brook, N.Y., 143p.
- (11) Considine and Frieswyk, 1982. (Review Draft) Forest Statistics for New York-1980, U.S. Forest Service, Northeastern Forest Experiment Station, Broomall, P.A.

- (12) Adirondack Park Agency, 1973. Adirondack Park Land Use and Development Plan, Adirondack Park Agency, Ray Brook, N.Y. 31p.
- (13) \_\_\_\_\_ Adirondack Park Agency Act, N.Y. Executive Law, Article 27 [First Set-out].
- (14) Adirondack Park Agency, 1973. Adirondack Park Land Use and Development Plan, Adirondack Park Agency, Ray Brook, N.Y., 31p.
- (15) Adirondack Park Agency, 1972. Adirondack Park State Land Master Plan, Adirondack Park Agency, Ray Brook, N.Y. 40p.

## INTERACTIVE MANAGEMENT AND UPDATING OF SPATIAL DATA BASES

P. French and M. Taylor  
Resources Planning Associates, Inc.  
P.O. Box 2562, East Hill  
Ithaca, N.Y. 14850 USA

### 1. INTRODUCTION

The decision making process, whether for power plant siting, load forecasting or energy resource planning, invariably involves a blend of analytical methods and judgment. Management decisions can be improved by the implementation of techniques which permit an increased comprehension of results from analytical models. Even where analytical procedures are not required, decisions can be aided by improving the methods used to examine spatially and temporally variant data.

This paper will discuss how the use of computer aided planning (CAP) programs, and the selection of a predominant data structure, can improve the decision making process.

### 2. SPATIAL DATA BASES - A NECESSITY

Modern society imposes a large number of constraints on the planning and management of those goods and services which affect the general welfare. One such example is the need to develop environmental impact statements for large scale construction projects. In order to assess future impacts which might result from a proposed project, a baseline condition first must be established. Data must be obtained to define the present state of the air, water, ecological and other natural resources, as well as the current economic, social and demographic characteristics. All of these types of information can be represented conveniently by spatial data bases.

In addition to presenting data as a baseline, some of these data also may be required as input to analytical models. These models will generate results (which in themselves may constitute a spatial data base) which, together with the baseline, can be used in the decision making process of project evaluation.

Defining the need for spatial data is easy. Obtaining all that is required can be difficult. Everyone collects and stores spatial data; the amount and level of sophistication vary from NASA's LANDSAT to the Department of Public Works supervisor who, after 35 years on the job, is the only one who knows the location of every storm sewer in the city. The complexity of most sources of data, however, lies somewhere between these two extremes, and invariably involves hardcopy maps on paper, Mylar, or linen.

The numerous sources of maps available create two problems: redundancy and accuracy. Since very few units of government have a Department of Cartography or other centralized spatial data base repository, it is incumbent upon each agency or department to both obtain and maintain the data it requires. Since jurisdictions (and sometimes functions) overlap, redundancy occurs. Due to differences in the frequency and level of updating, this redundancy eventually leads to discrepancies. Therefore, in addition to the problem of where to go to obtain data, one also may have to contend with the problem of which source to believe.

Problems associated with the suitability and accuracy of necessary data can be reduced through the careful development, management and updating of comprehensive data bases. Although every planner and manager may dream of the perfect data base - one that has every conceivable type of information, 100% accurate, and up to date - this "wealth" of information could prove valueless if the decision maker is unable to manipulate it, comprehend it, or assess its significance. As the volume of accumulated data increases, it becomes increasingly difficult to maintain and rely upon a system based solely upon paper maps.

One method found to be extremely useful in assisting decision makers in understanding data complexities and interactions relies upon the use of interactive color computer graphics to both help generate and display complex spatial data bases. While the ability to quickly and easily analyze and interact with spatial data is important, no less important is the display format selected for presenting the digital data. Developers of digital cartographic systems should attempt to provide a data base that, when displayed, looks as good as a high quality paper map.

### 3. SPATIAL DATA BASE STRUCTURES

In order to provide a data base which is easy to manipulate and comprehend, the structure selected for the data base is important. The conventional structures involve points, lines, areas or, as a simplification of a polygonal structure, grid cells. The different relationships that can exist between geographical entities, and the various searching and sorting operations which can be encountered, have been discussed by Nagy and Wagle (1979). Rather than reviewing those discussions, some of the advantages and disadvantages associated with each type of structure will be presented.

#### 3.1 Point Data Structures

Perhaps the greatest advantage to using a point data structure is the relatively small amount of storage required. However, exclusive use of a point system limits the information portrayed to discrete locational data: utility poles, manholes, geologic core samples, well locations. For certain types of point data (such as elevation or depth to groundwater) even a sparse data set can be sufficient for determining probable



values at unknown locations if appropriate techniques such as Kriegering (Delfiner and Delhomme, 1975) are used. Unfortunately, if connectivity between points exists, it is not retained in a point data structure. This contributes to the difficulty of trying to geo-reference the points within the area of interest.

### 3.2 Line Data Structures

As with points, the use of line primitives is relatively low cost from the standpoint of storage and processing requirements. If linear representations of curvilinear segments are used, the storage and processing requirements are further reduced. Also, use of linearized primitives can provide a de facto point data structure, as long as the points are associated with the particular network (transmission towers, substations, bus stops).

While line data structures provide more information than point structures from the standpoint of visual geo-referencing, a line or line-point data structure is inadequate for representing most geographical data bases.

### 3.3 Area Data Structures

The use of polygonal or area data structures literally provides a new dimension in the ability to portray information, since zones of an attribute type or value now can be described and quantified. If the polygon boundaries are stored as linear line segments, along with the two attribute types they separate (thus preventing duplication of boundary definition for adjacent polygons), the storage requirements can be modest if the area/perimeter ratio is large.

From the standpoint of visual geo-referencing, polygonal data structures are less informative than linear data structures when viewed apart from everything else. Their utility is not apparent until viewed in the proper spatial context: a utility service area displayed with the transmission network, or as an overlay on an appropriately scaled map. When combined with other data structures, polygonal data can be very informative. However, while the ability to answer a question such as "Locate and quantify the miles of transmission lines within a number of service districts which provide areas not meeting certain population density requirements" becomes possible, arriving at the answer can be computationally intensive.

### 3.4 Grid Cell Data Structures

Given a fixed grid system, each grid cell occupies a known geographical location, has a known area, and can have multiple attributes associated with it. These characteristics facilitate the overlaying of and

determination of intersecting areas of different attributes. This type of data structure can be interacted with and updated easily and, as will be shown, can provide the visual geo-referencing necessary to make the system user friendly.

Another advantage of a grid cell based data structure is the link that can be made with video technology. Grid cell data not only can be displayed on video monitors, but also can be created using video equipment.

The disadvantages of using a grid cell system primarily arise from the size of the cell. The smaller the cell size, the greater the ability to approximate a point-line-area data structure. However, if the cell size is too small, data input methods and storage requirements become critical. Indeed, manual definition of cell data may be too labor intensive and prone to error for very small cell structures. Conversely, data for large cell areas become easier to input and store, but at the expense of detail and accuracy: points and lines no longer are represented realistically.

#### 4. VIDEO BASE MAPS: A PREDOMINANT DATA STRUCTURE

Certainly no single data base structure provides all of the advantages and none of the disadvantages associated with a geographic information system. Where man-made attributes have to be considered, point and network data become commonplace. For analyses required for site planning purposes, the use of some type of polygonal data is virtually assured. Given that different data structures are required to fully represent the requirements of geographic information systems, consideration should be given to the selection and creation of a predominant data base structure. The selection should be one which provides for the best visual quality of the displayed data, enhances the utility of the other data base structures, and is easily updated.

In our work with water and other natural resources planning, we have found that the use of computerized video base maps satisfies our criteria for a predominant data base structure. The computerized grid cell base maps, created by video digitizing aerial photography, U.S.G.S. 7 1/2 minute quad sheets, or other suitable hardcopy, contain the complex "extras" - the highschool track, the interstate's cloverleaf, parking lots, fields under cultivation - which provide for instant recognition and visual geo-referencing. If additional data to describe the location of points or networks are required, they should be contained in, and extracted from, supplemental point-line data bases and registered onto the video base map. In this manner, the video base map enhances the utility of the supplemental data through improved comprehension of spatial relationships.

The use of grid cells as the predominant data structure, and color computer graphics as the display medium, enable utilization of the many

private and Federal data bases which are becoming digital. Examples include worldwide LANDSAT coverage, digital terrain models which are being developed for the entire U.S., and Census data. All of these data bases, which are grid cell, point, line or polygon based, can be geo-referenced and displayed over the video base maps.

In addition to improving a person's ability to mentally geo-reference data, the use of video technology also permits data to be displayed in color. Some geographic information systems which utilize color have only destructive color displays. That is, when displaying color-coded polygonal information, any underlying data are masked by the addition of color. Since destructive displays also inhibit the ability to geo-reference data, non-destructive display techniques should be used whenever possible. These techniques have been developed and incorporated into a computer-aided resources planning package by the authors. In several test applications this package has demonstrated the value of these techniques in the planning process. Figures 1 and 2, illustrating typical displays from this package, are based upon the Rochester East and Webster, NY 7 1/2 minute quadrangle maps.

## 5. CREATING/UPDATING SPATIAL DATA

The ability to define and update complex, large scale spatial data bases also is facilitated by the use of video techniques. Original color (or black and white) hardcopy maps or photographs can be quickly converted to a digital grid cell structure. Since the data can be defined at up to video rates (1/30 second for a 525 line image), large amounts of information can be captured quickly in a non-labor intensive manner. Supervised classification techniques then can be used to convert three channel (red, green, blue) data to a single classified channel. With all pertinent spatial information available in a single channel, the ability to rapidly interact with the data is improved.

Interactive updating of video maps can be performed using any of several methods. Point, line or polygon data can be entered using a digitizing pen and tablet and locating the position visually, or by registering a source map on the tablet to the video image and tracing the appropriate information. Updates also can be made using the many available sources of digital information which have been suitably classified and registered to the video base map.

## 6. HARDWARE AND COST

There are many people who will dismiss the use of grid cell digital cartographic data bases either for the reason mentioned in section 3.4 (too large a storage requirement), or because of a lack of resolution. Concern over the latter point probably is justified with the use of commonplace raster graphics hardware, which has 480 x 512 displayable cells and provides a resolution of approximately 0.046" for a 7 1/2

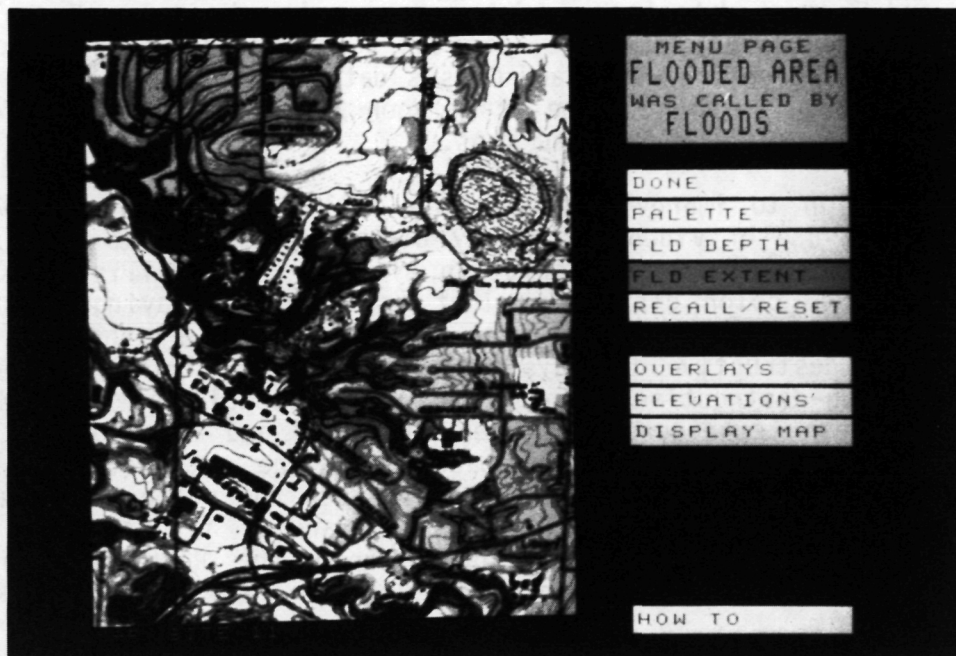


Fig. 1. Results from a flood analysis package are displayed (blue) over the 1 channel (8 bit) video base map.



Fig. 2. Color-coded land use information, derived from polygonal data, are registered and displayed over the 1 channel video base map.

minute quad sheet. However, hardware capable of displaying 1000 x 1000 cells are readily available; units that display 2000 x 2000 cells, each with 8 or more bits of color, will be available in the next 3-5 years. A device with 2000 x 2000 resolution corresponds to a quad sheet resolution of approximately 0.01 inches/cell, considerably better than the National Map Accuracy Standards (Monmonier, 1977).

The concern over the vast amounts of data storage required for grid cell maps also will subside in the next three to five years. As computer-compatible read/write video disc equipment becomes available, digitized maps (areal photographs, terrain models, etc.) will be stored and retrieved quickly. Even using today's video standards, a 2000 x 2000 cell composited image could be displayed in approximately 1/2 second. At this resolution, digital coverage for each of the more than 40,000 7 1/2 minute quadrangles in the U.S. could be stored on 3 movie length video discs.

Equipment needed to provide the interactive capabilities mentioned can be obtained for as little as \$35,000. While this price currently may be exclusionary for some of those groups which most utilize spatial data - local, county, regional and state planning agencies - the cost of computer hardware continues to decrease. Also to be considered are the cost and availability of the necessary software. Although the cost of software could approach that of hardware, we foresee a number of micro-computer-based turnkey video cartographic data systems being offered for less than \$40,000 in the next 3 to 5 years.

## 7. CONCLUSIONS

The ability to visually geo-reference complex spatial data and spatially oriented results from mathematical models is an important characteristic required of many planning processes. The use of a predominant grid cell data structure, based upon the use of video images and color computer graphics, helps to provide the geo-referencing capability. This data structure and the computer and video hardware facilitate interactive creation, management and updating of complex spatial data. The ability to rapidly manipulate large amounts of spatial data, perform analyses with that data, and display color coded results which can be visually geo-referenced to that data, help to place in proper perspective the role of judgment in the decision making process.

## 8. ACKNOWLEDGEMENTS

The computer-aided planning package used to generate Figures 1 and 2 was developed at Cornell University by the authors under the direction of D.P. Loucks; funding was provided by a grant from the National Science Foundation.

## REFERENCES

Delfiner, P. and Delhomme, J.P. "Optimum Interpolation by Krieking." In Display and Analysis of Spatial Data, J.C. Davis and M.J. McCulloch (Eds.), John Wiley & Sons, Inc., New York, 1975.

Monmonier, M.S. Maps, Distortion, and Meaning. RP 75-4, Association of American Geographers, Washington, D.C., 1977.

Nagy, G. and Wagle, S. "Geographic Data Processing." ACM Computing Surveys, V. 11, No. 2, June, 1979, pp. 139-181.

## THE OKLAHOMA GEOGRAPHIC INFORMATION RETRIEVAL SYSTEM

W. ANTHONY BLANCHARD

Center for Applications of Remote Sensing

Oklahoma State University, Stillwater, Oklahoma, U.S.A.

### ABSTRACT

The Oklahoma Geographic Information Retrieval System (OGIRS) is a highly interactive data entry, storage, manipulation, and display software system for use with geographically referenced data. Although originally developed for a project concerned with coal strip mine reclamation, OGIRS is capable of handling any geographically referenced data for a variety of natural resource management applications. A special effort has been made to integrate remotely sensed data into the information system. The timeliness and synoptic coverage of satellite data are particularly useful attributes for inclusion into the geographic information system.

### INTRODUCTION

The Center for Applications of Remote Sensing (CARS) at Oklahoma State University performs a major public service role by providing Oklahoma with new technologies in a wide range of disciplines. Computer based geographic information systems have become an important area of interest at CARS because of their features which allow enormous quantities of data to be managed quickly and efficiently [1]. The Oklahoma Geographic Information Retrieval System (OGIRS) was developed at CARS in response to a request by a state agency (the Oklahoma Department of Mines) for a geographic information system which could integrate remotely sensed data with field and archive data sources.

The basic concepts and techniques utilized in the OGIRS software package do not contain any outstanding advances in the state of the art of computer based geographic information system methodologies. The main criteria for the design of OGIRS were simplicity, flexibility, and efficient operation on a small computer, therefore, many proven techniques were employed in the design [2,3]. The program is very simple to use and little or no previous computer ability is required by users. Requests and system prompts are in plain English or a simple three letter mnemonic code. OGIRS accepts data in digital form from previously established magnetic disk files, or the program provides a digitization module for use with either an on-line graphics digitizer or as input from a remote terminal. All data are stored in a common format and are geographically referenced to the Universal Transverse Mercator (UTM) grid [4]. The program structure utilizes the modular overlay capabilities of the Perkin-Elmer 8/32 host minicomputer. OGIRS requires approximately 50 percent of the minicomputer's available core memory of 0.5 megabyte at any given time. The program accesses only those data files and devices required for immediate execution to save time and space.

### BACKGROUND

Actual programming of the Oklahoma Geographic Information Retrieval System began

in late 1981, however, the concept began taking shape as early as 1976. A project funded by the United States Department of the Interior was undertaken by the Louisiana State University Division of Engineering Research to monitor and assess energy related activity impacts to the water resources of south Louisiana [5]. The work on this project brought into focus many of the needs currently addressed by the Oklahoma Geographic Information Retrieval System. The USDI project called for the development of methodologies which would locate geographic areas most susceptible to environmental degradation from energy related activities (primarily oil and gas exploration, transportation, and production). A technique developed by the Battelle Laboratory of Columbus, Ohio [6] was chosen as the basic model to be further elaborated on by the LSU project. The final LSU suitability analysis methodology required cellularization of the study area, a general description of the physical environment of each cell, and a complex weighting system for evaluating an environmental impacts matrix for each land cover category.

Remotely sensed data were employed in the LSU study on a cursory basis to map the vegetation of the study area. A computer was used to reduce the time required to establish weighted values from each environmental impacts matrix, and a geographic information system was developed from the land cover and physical environment data. All of the prerequisites for a computer based geographic information system were compiled during the LSU study, however, hard copy maps and clear acetate overlays were used as storage media instead of a computer data base. It became evident during the LSU study that automation, a more efficient data storage medium, and more flexible data manipulation methods were necessities when an attempt is made to manage large amounts of geographically referenced information. Specifically, the need to incorporate data from a variety of sources and in a range of original formats, especially remotely sensed data, and the ability to interactively design and implement applications models were prime motivations for the subsequent development of OGIRS.

The Center for Applications of Remote Sensing began a project funded by the Oklahoma Department of Mines in 1981 to develop techniques for monitoring reclamation of coal strip mines [7]. A preliminary examination of the problem revealed many of the same needs and difficulties evident in the 1976 LSU/USDI project. The Oklahoma Geographic Information Retrieval System began as a simple computer program designed to read coordinates from a graphics digitizer and load them onto magnetic tape. The program quickly expanded and now functions as a fully operational data base program for geographically referenced data. OGIRS is still a developmental program with new capabilities added on a continuing basis, however, a stable production version of the software is in residence on the CARS computer.

In the near future, OGIRS will have a color image display capability and a polygon to raster conversion module of its own. A continual upgrading and enhancement of the basic software package is planned, but no major changes to the program format are in the offing. The largest single addition to the program will be a predictive land cover modeling (PLCM) overlay to be added sometime in 1983. The PLCM will allow a user to interactively develop possible future scenarios by altering one or more of the existing physical parameters in a defined geographic area. The program will graphically display the possible changes in the land cover through time.

A microcomputer version of OGIRS is under development at CARS. Although the major features of the micro-OGIRS will be identical to the minicomputer version, the data handling and display capabilities will be smaller than in the original version. The anticipated date of completion for the micro-OGIRS program is mid 1983.

## PROGRAM DESCRIPTION

In general, the Oklahoma Geographic Information Retrieval System has many features found on similar software systems available today [8,9]. The data entry, stor-



age, and display techniques are conventional in most respects, however, these functions have been optimized for the specific operating environment. The data manipulation algorithms are also common to many other geographic information systems [10]. The unique feature of OGIRS is the flexibility users have in constructing applications models interactively.

OGIRS is programmed in FORTRAN VII and is currently implemented on a Perkin-Elmer 8/32 host minicomputer equipped with the OS32 version 4.3 operating system. The minicomputer is the primary image processing system at the OSU Center for Applications of Remote Sensing. The CARS computer has a number of specialized peripherals that may not be commonly available, such as the digital image display system. The CARS system also has a graphics digitizer, an electrostatic printer/plotter, and a color product generation system which are not required, but are very desirable options.

The program structure consists of a root program and a series of overlay programs (see Figure 1). OGIRS uses Perkin-Elmer operating system dependent subroutines for dynamic allocation and assignment of logical units and disk files [11]. The portability of OGIRS has not been established at the present time.

Data entry and storage are accomplished by several means. OGIRS interfaces with a graphics digitizer and records x, y, and z coordinate values from a cellularized data set. New z values may be interactively entered from a remote terminal once an x and y coordinate set is established for a given study area. Scale, cell size, and coordinate system are user selectable options.

OGIRS interfaces with the NASA Earth Resources Laboratory Applications Software System (ELAS) [12] and makes use of the ELAS polygon to raster conversion algorithm for input of polygon data sets. OGIRS also makes use of the digital image display capabilities in the ELAS program.

Data are stored as digitized data sets collected into specially constructed magnetic disk files or thematic libraries. A thematic library is titled after the major theme of the data entered in it. Generally, five basic libraries (soil, geology, hydrology, climate, and land cover) are sufficient to include most data types. Each thematic library is divided into channels or individual data sets representing some specific geographically referenced information (see Figure 2).

Figure 1. OGIRS program structure.

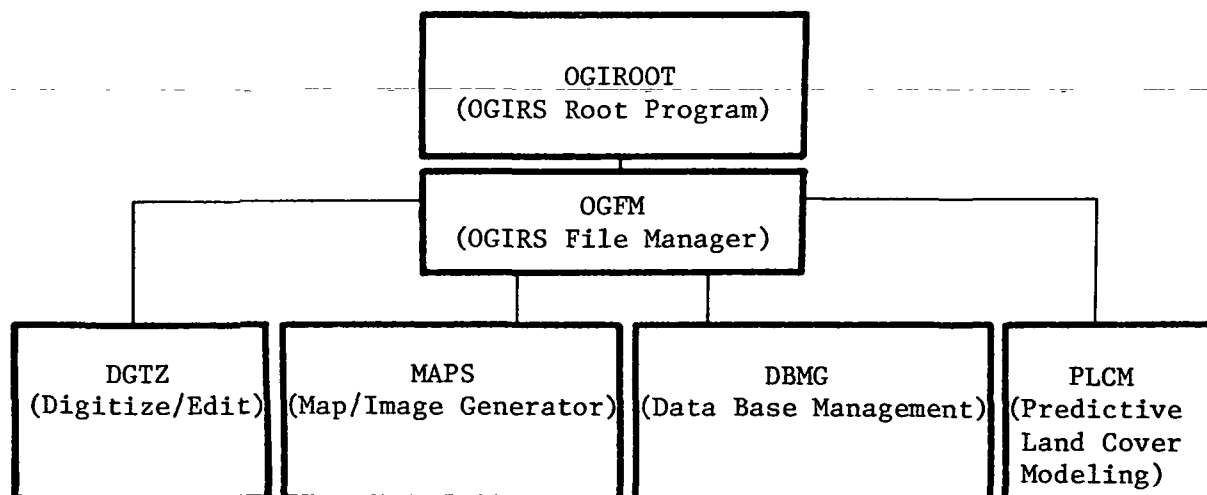


Figure 2. OGIRS thematic library contents.

GEOLOGIC LIBRARY	LANDCOVER LIBRARY	SOILS LIBRARY	TOPOGRAPHIC LIBRARY	CLIMATOLOGICAL LIBRARY
1. Surface geology	1. Landcover	1. Soil series	1. Slope angle	1. Mean annual precipitation
2. Depth to bedrock		2. Shrink/swell potential	2. Slope aspect	2. Mean January temperature
3. Type of bedrock		3. Soils subject to floods	3. Elevation	3. Mean July temperature
		4. Soil permeability		4. Mean length of frost-free season
		5. Hydrologic soil group		
		6. Soil Erodibility		

## 1. OGIRS PROGRAM OVERLAYS

Each of the OGIRS program overlays functions as a separate task within the Perkin-Elmer operating environment. Any overlay may be entered from any other overlay, therefore, there is not a required program flow. Upon entering the program, the master file is accessed (if a master file does not exist, the program creates one) and the devices and thematic library files are made ready. The user is always located in the OGFM overlay when program execution begins.

### 1.1 OGFM

The OGFM overlay is the OGIRS file management program. This overlay allows the user to allocate indexed or contiguous files, assign disk files or peripheral devices as logical units, deassign files or devices, list logical units and their status, and list the names of other overlays available. As with all of the OGIRS overlays, OGFM provides a list of directives and is highly interactive. A user types the appropriate three letter mnemonic code to select any of the program capabilities.

Disk files and peripheral devices are designated for specific uses when they are assigned. OGFM has four designations for a file or device: input, output, print, or null. The designated usage may be changed interactively.

### 1.2 DGTZ

The DGTZ overlay is the OGIRS data entry and edit program. As previously mentioned, data are entered from a graphics digitizer, a remote terminal, or as digital image data sets. The DGTZ edit functions allow the user to add, delete, or change the data value of any element within the input data set to correct for missing, superfluous, or incorrect values. A no-edit option is available for data sets with no er-

rors. Data are reformatted during the edit or no-edit runs and loaded into a thematic library. A library update function is available as a post-reformat edit capability. DGTZ contains a point listing capability for hard copy output of data sets in tabular form. A directives list for DGTZ is provided on entry to the program.

### 1.3 MAPS

The MAPS overlay is the OGIRS map and digital image generation program. This module outputs to a line printer or remote terminal an ASCII character set plot of the selected thematic library channel or DBMG generated data set. These products, although functional are primarily intended to be used as previews of the color image displays and electrostatic printer/plotter maps. In addition to the actual map of the data values, the header information, a frequency distribution, a value/character legend, and a map summary including the total number of categories, the total number of elements, and the area are printed with each map (see Figure 3).

Digital images designed for display on the CARS digital image display system are produced with the IMG directive in MAPS. The resulting images (see Figure 4) are compatible with the NASA Earth Resources Laboratory Applications Software System (ELAS). Elas is used to display the MAPS generated images at the present time, however, an image display overlay for OGIRS is under development. A directives list for MAPS is provided on entry to the program.

### 1.4 DBMG

The DBMG overlay is the OGIRS data base management program. Both data manipulation of single data sets and interactions between data sets are available as user selectable functions in DBMG. Three categories of data base management functions are available: arithmetic functions, relational modes, and mapping functions. A directives list for DBMG is provided upon entry to the program.

The DBMG arithmetic functions include the following set of operations: add, multiply, divide, square root, exponentiation, and natural logarithm. These operations may be applied to a thematic library data set and a constant, or between two or more data sets. The resulting data set is stored in the OGIRS master file until it is saved in a thematic library.

The five DBMG relational modes include: equals, greater than, less than, greater than and less than, and less than or greater than. Relational modes operate on individual data sets and user selected constants. The modes are used as a feature selection function giving the user the ability to isolate values, or ranges of values, from the total data set.

DBMG's mapping functions allow a user to construct composite maps from two or more thematic library data sets. Composite data sets may be generated using intersection, union, or exclusion set theory functions. The resulting data set is stored in the master file until it is saved in a thematic library.

Each of the three data base management functions has a valuable utility for manipulating data when used separately, however, when one or more of the functions are used consecutively their power increases. The interactive capability of the DBMG program overlay provides the user with a flexible model building tool. Very few constraints are placed on the user allowing the range of potential models to extend from computing soil loss to site suitability analysis. Any algorithm which uses the arithmetic functions, relational modes, and mapping functions described above may be interactively implemented on all or any part of the data available in a thematic library.

Figure 3. Line printer map produced by the MAPS overlay.

EASTMINE HYDROLOGIC SOIL GROUP  
 CRAIG  
 IL = 1 LL = 10  
 IE = 1 LE = 13  
 CH = 4  
 DATE: 25/02/82 LIBRARY DESCRIPTOR: SOIL

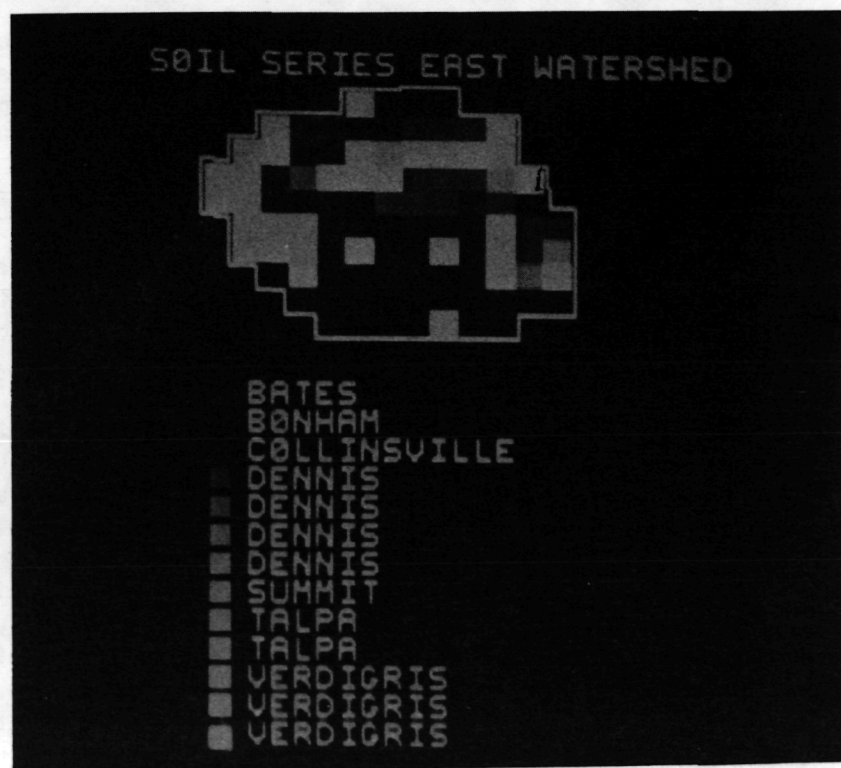
```

      A B B C
    A B B B B B B C
  A A B B C C C C C C
A A B B C C C B B B C
A A B B C C C B B B B
  A A A B C C B B B C B B
  A A A B C C B C B C B B
    B A B B B B B C B C
      B B B B B B B B
        B B B A B B
  
```

CAT #	VALUE	SYMBOL	FREQ
1	4	A	16
2	3	B	63
3	2	C	20

3 CATEGORIES 99 CELLS 990A AREA

Figure 4. Digital image generated by the MAPS overlay.



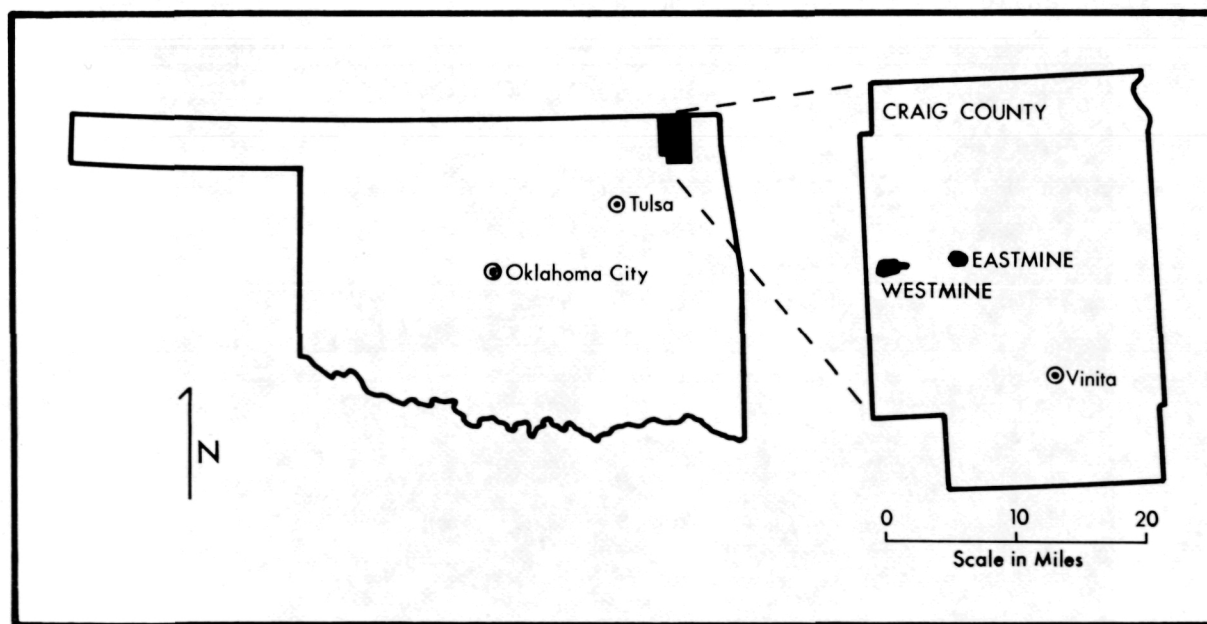
## APPLICATIONS

The Oklahoma Geographic Information Retrieval System was developed for a pilot project concerned with geographic information system applications to surface mine management [13]. The Oklahoma Department of Mines provided the funds for the initial project. The main objectives of the study were: 1) to provide a computer based geographic information system to store and retrieve many disparate data sources including remotely sensed data from satellites, field data, and available map data; 2) to provide techniques for manipulation of the data sources; 3) to provide mine inspectors in the field with the data reduced to a manageable, easily interpreted format to determine compliance with state and federal mined land reclamation laws.

Oklahoma's Mining Lands Reclamation Acts of 1968 and 1971 [14] and the United States Surface Mining Control and Reclamation Act (PL 95-87) [15] require all mine operators to reclaim strip mine lands. The need for effective reclamation is keenly felt in Oklahoma where 14326 ha (35400 acres) of land had been disturbed by surface mining by 1973. Less than 25 percent of the total mined land had been reclaimed by 1973 [16]. A resurgence in surface mining occurred in the post energy crisis years of the 1970's and the demands on inspectors to cover the greatly expanded area and number of mines led to the need for updating inspection techniques. The choice was to use satellite (Landsat) remotely sensed data to monitor conditions at mine sites, however, the satellite data were capable of providing only a few of the many variables inspectors had to contend with when monitoring mine sites. A geographic information system which incorporated the timely land cover information derived from Landsat, archived geological, hydrological, pedological, climatological, and other natural resource data, and field site inspection information became a realistic solution to a mammoth data management problem.

Two small watersheds in northeastern Oklahoma's Craig county, 88.6 km (55 miles) northeast of Tulsa (see Figure 5) were chosen as study areas. The eastern watershed covers 978 ha (2416.6 acres) and includes an active coal mine. The western watershed

Figure 5. Location of study areas.



is 1403 ha (3466.8 acres) in size and does not contain any mining activity at the present time.

Data for both watersheds were collected, digitized, and stored in thematic libraries. Figure 2 presents the contents of the thematic libraries. The flexible data manipulation capabilities of the OGIRS software package were used to generate new data sets from the disparate data in the thematic libraries. Examples of the data presentation, feature selection, and modeling capabilities of OGIRS follow.

The archived data sets were used unchanged as presented as hard copy output maps to supply needed information on the physical environment of the study sites. A slope angle map is shown in Figure 6. The map was used to determine critical slope areas as well as baseline slope information.

The feature selection capability of OGIRS was used to isolate specific features from complete data sets and then present the extracted information as a new map (see Figure 7). Specific soil types were isolated from the soil series data set by this method. In addition, the OGIRS mapping functions were used to combine two or more data sets generated by the feature selection mode. A map showing areas with the greatest runoff potential was constructed by intersecting the runoff potential map with a map representing only the steepest slope angles (see Figure 8).

The modeling capability of OGIRS is specifically designed to allow maximum user flexibility. Consequently, few structured modeling modules have been included in the program package. The coefficient of areal association (CAA) model has been included as a module because of its importance in displaying the areal correspondence between two maps. The algorithm for computing the CAA [17] is included in the MAPS overlay. The CAA is a statistical model that describes the areal correspondence of two maps as a decimal value between 0 and 1. A zero value describes total dissimilarity, whereas a value of one indicates the two maps are spatially identical. The CAA model was used to determine which types of bedrock are most commonly associated with high elevations. A map displaying a specific bedrock type selected from the bedrock data set in the geological thematic library is compared to the map displaying elevations over 243.8 m (800 feet). The resulting CAA value of 0.4349 gives the degree of areal association of that bedrock type to high elevations..

Figure 6. Slope angle map for the eastern watershed.

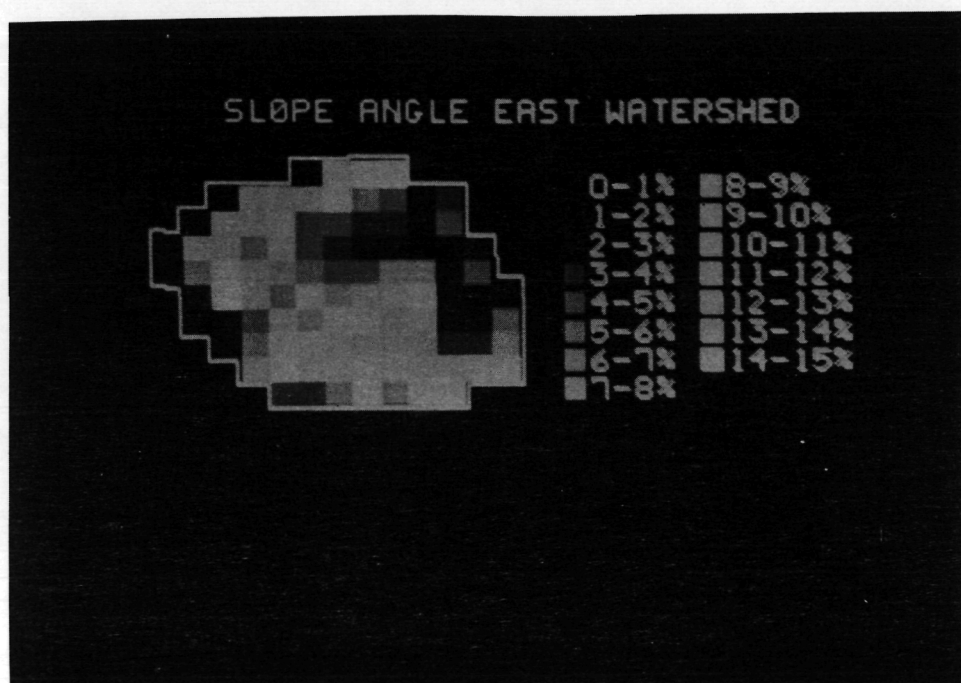
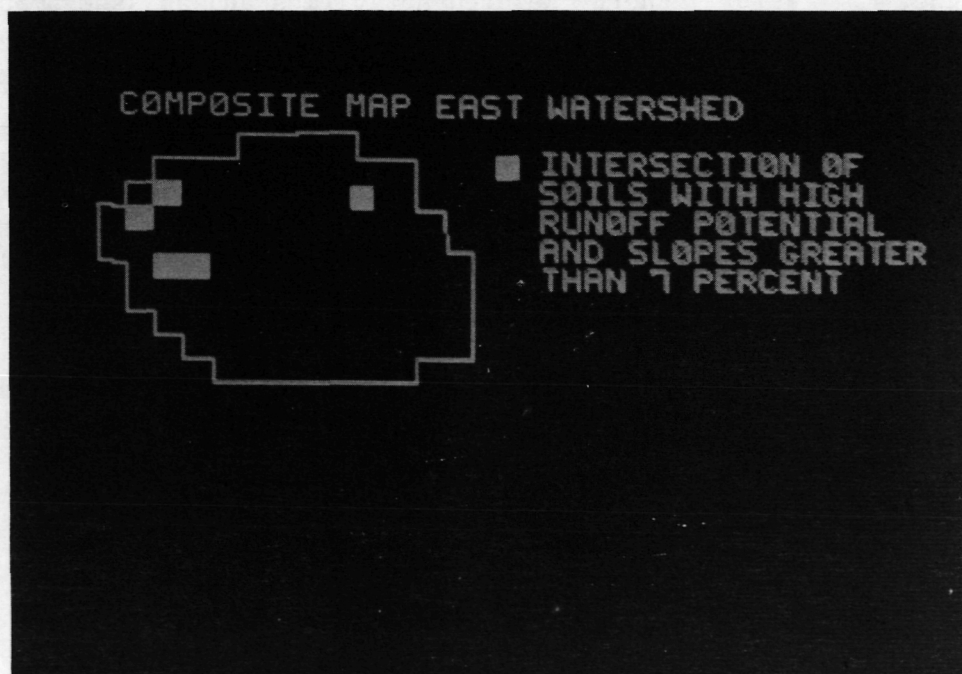




Figure 7. Map displaying feature selection capability.



Figure 8. Greatest runoff potential map.



## CONCLUSIONS

The Oklahoma Geographic Information Retrieval System (OGIRS) is a flexible, highly interactive data entry, storage, manipulation, and display software system for use with geographically referenced data. The program is designed for simplicity and efficient operation. Data are entered from a variety of original sources including: remotely sensed digital data, field samples, and archived information. The digitized data are stored in a common format in thematic libraries on magnetic disk. OGIRS allows a user to design applications models from the arithmetic operands, relational modes, and set theory mapping functions available in the data base management overlay.

The OGIRS software package has been applied to surface coal mine management and reclamation for a small pilot project in Oklahoma. The program allowed a large quantity of information to be managed quickly and efficiently.

## REFERENCES

1. Cicone, R. C. 1977. Remote Sensing and Geographically Based Information Systems. Proc. of the Eleventh International Symposium on Remote Sensing of the Environment. 2:1127-1137.
2. Wilson, C. L., and Thomson, F. J. 1981. Integration and Manipulation of Remotely Sensed and Other Data in Geographic Information Systems. Proc. of the Pecora VII Symposium. 303-317.
3. Smith, A. Y., and Blackwell, R. J. 1980. Development of an Information Data Base for Watershed Monitoring. Photogrammetric Engineering and Remote Sensing. 46:8:1027-1038.
4. Raisz, E. 1962. Principles of Cartography. McGraw Hill. New York. 315 pp.
5. Whitehurst, C. A., Kinney, R. A., Muller, R. H., Milliet, M. P., Blanchard, W. A., and Zeringue, F. 1976. Energy Related Activities and an Assessment of the Water Resource Management Alternatives in South Louisiana. LSU Press. Baton Rouge. 270 pp.
6. McGinnis, J. T. 1975. Environmental Aspects of Gas Pipeline Operations in the Louisiana Coastal Marshes. Final Report Battelle Columbus Laboratory. Columbus.
7. Blanchard, W. A., Bunse, L. M., and Walsh, S. J. 1982. Pilot Study: Oklahoma Geographic Information Retrieval System Applications to Surface Coal Mine Management. CARS. Stillwater. 79 pp.
8. Amidon, E. L., and Dye, E. J. 1981. SCANIT: Centralized Digitizing of Forest Resource Maps or Photographs. USDA Forest Service General Technical Report PSW-53. Berkeley. 42 pp.
9. Mitchell, W. B., Guptill, S. C., Anderson, K. E., Fegeas, R. G., and Hallam, C. A. 1977. GIRAS: A Geographic Information Retrieval and Analysis System for Handling Land Use and Land Cover Data. USGPO. Washington.
10. Dobson, J. E. 1979. A Regional Screening Procedure for Land Use Suitability Analysis. Geographical Review. 69:2:224-234.
11. Perkin-Elmer. 1979. FORTRAN VII Run Time Library Reference Manual. Publication No. S29-578R03. Oceanport.



12. Junkin, B. G., Pearson, R. W., Seyfarth, B. R., Kalcic, M. T., and Graham, M. H. 1980. ELAS: Earth Resources Laboratory Applications Software System. NASA National Space Technology Laboratory. Bay St. Louis.
13. Bunse, L. M., Walsh, S. J., and Blanchard, W. A. 1982. Surface Coal Mine Management and Assessment Through A Remote Sensing Based Information System. Proc. of the Annual Meeting of the American Association of Geographers. in press.
14. \_\_\_\_\_. 1981. The Mining Lands Reclamation Act 1971. Oklahoma Statutes 1981 Courts - Professions and Occupations. 2:3222-3227.
15. \_\_\_\_\_. 1978. Surface Mining Control and Reclamation Act (PL 95-87). U. S. Code Congressional and Administrative News. 1:445-532.
16. Johnson, K. S. 1974. Maps and Descriptions of Disturbed and Reclaimed Surface Mined Coal Lands in Eastern Oklahoma. Oklahoma Geology Survey. Norman. 12 pp.
17. Taylor, P. J. 1977. Quantitative Methods in Geography. Houghton Mifflin Company. Boston. 386 pp.

DATA BASE MANAGEMENT  
FOR  
GEOGRAPHIC INFORMATION SYSTEMS

BY: Michael G. Pavlides  
Greenhorne & O'Mara, Inc.

Remote Sensing has permitted scientists to view "old" data - or data that have been examined previously by other techniques - from a variety of new and different perspectives. It has also permitted scientists to view and interpret "new" data - or data that have not previously been acquired and analyzed. The analytical processes involved (e.g., spectral signature analyses), as well as the technology itself from a data acquisition point of view are data comprehensive.

The technology of Geographic Information Systems (GISs) has automated the once manual technique of overlaying data for analysis and interpretation. When manual techniques were employed only the most important data overlays (or "data bases") were used since analysts could not possibly assimilate all available data or review all possible data combinations due to the traditional constraints of time, budget, accuracy and human factors. In other words, when performing overlay analyses manually, analysts were required, out of necessity, to be "selective" and choose only those data bases that provided the most comprehensive amount of relevant information for the intended objectives of the analyses. With the advent of increased data storage, computing speed, assimilation capacity, and analytical methods now available in GISs, a shift in this philosophy has occurred. The emphasis now is to add more data to required analyses in order to refine and make the results more accurate. Selective use of data is no longer the *modus operandi*. This use of GISs is also data comprehensive.

The integration of remotely sensed data with GISs, particularly in the field of energy resource management, has consequently resulted in the formidable problem of managing an extraordinary amount of data. "Data Base Management" (DBM) has become a generally accepted phrase utilized to describe a variety of data storage and manipulation functions, and, in the context of this paper, DBM refers to the "planned" management" or more specifically, the procedure for "order of entry" of the massive variety and quantity of data bases into Geographic Information Systems. A logical approach to determine the data base order of entry is presented herein utilizing management techniques and Consideration Factors (CFs).

Insight into the problem of the magnitude and variety of data bases as related to GISs can be further derived from a discussion of a classification of GISs. In an editorial featured in the July/August 1982 *Computer Graphics News* entitled "The Future of Geographic Information Systems", Dr. Robert Aangeenbrug classified five (5) types of GISs:

- GIS #1. Natural Resources Inventory Systems: Used to perform monitoring and evaluation functions of resource data (e.g., overlays of ecologically-sensitive habitats) for regulation of regionwide activities.

- GIS #2. Urban Systems: Serves a dual function as a Land Record Management file (for tax purposes, for example) and a related engineering design data file (e.g., topographic data).
  
- GIS #3. Planning And Evaluation Systems: Used generally to provide thematic display of the entire realm of geographic data (such as socioeconomic information) most frequently by general or relative spatial relationships for use by planners and policymakers (Dr. Aangeenbrug cites the Decision Information Display System as an example).
  
- GIS #4. Management, Command And Control Systems: Used for "strategic" planning determinations by industry and military planners. According to Dr. Aangeenbrug this GIS is similar to GIS #3 with the exception of actual program structure. However, another difference lies in the analytical inference capabilities of these systems to derive systematic relationships from examination of combinations of data bases. Whereas GIS #3 may display the number of unemployed individuals by county in a state, GIS #4 may relate the unemployed individuals to their previous annual income to allow the analysts to conclude what level of jobs are being lost with most frequency.
  
- GIS #5. Citizen/Scientist Systems: Provides the user, access to informational data bases through common telecommunication carriers such as home computers and television sets.

Dr. Aangeenbrug has, of course, developed a broad categorization of GISs as a function of applications rather than internal program structure. It is interesting that, broadly speaking, each of these application functions represent a relatively distinct type of data set.

GIS #1 deals with scientific data or the physical and/or environmental (i.e., "real") characteristics of eco-terrain units (e.g., the polygonal area of a specific soil type). Since the boundaries of most eco-terrain unit data are interpreted and subject to natural, ongoing change rather than "absolutely fixed" in space or time, relative (rather than true geographic) spatial relationships between data subsets are ordinarily sufficient for purposes of GIS analyses. GIS #2 deals with "engineered or measured" data, which are important from an accurate (rather than relative) geodetic spatial relationship (e.g., property tax maps with meets and bounds, planimetric maps based on state coordinate grid systems). GIS #3 deals with "descriptive data" relating general information of the geounits, most often social and economic characteristics (e.g., the number of voters in a county). GIS #4 deals with developing a data base of "data interrelationships". Although not stated by Dr. Aangeenbrug, this GIS directly, or indirectly through user interpretation, is used to derive "interrelationship data bases" that are directly relative to intended user objectives, from analyses of inputted data bases (e.g., development of income statistics related to levels of



energy consumption in a county). GIS #5 deals purely with "informational data" or the display of data that may or may not be geounit-oriented. GIS #5, by definition, might include all data in GIS #1 through GIS #4 and any other information of a general interest to potential users (e.g., the stock market history of a given corporation, the number of cancer patients in a county). In summary:

<u>GIS #</u>	<u>System Name</u>	<u>Data Set Type</u>
1	Natural Resource Inventory	Real
2	Urban	Measured
3	Planning & Evaluation	Descriptive
4	Management, Command & Control	Interrelational
5	Citizen/Scientist	Informational

All of these data set types are further subdivided into data subsets. Subsets for the "Real" data set type are, for example, structural geology, fault and folds, soil types, vegetation and so on. If we consider each possible data set type and subset as a possible data base, the enormity of the data entry problem becomes self-evident.

There is no doubt that GISs exist, or are in process of development with program structures capable of handling multiple GIS application functions and the corresponding data set types inferred by those applications. With the advent of these more sophisticated GISs that are capable of handling more data sets and subsets, the GIS manager is confronted with the significant problem of ordering data base entry to the GIS. This is further complicated by (1) limitations of digitizing budgets, (2) primary user requirements, and (3) by traditional system management problems (monotony of digitizing causing quality problems, high personnel turnover, and so forth).

GIS managers have attempted to solve this problem by immediate entry of data bases required to satisfy user needs without affording the GIS data management concept a more holistic approach. The objective of data base entry ordering is to (1) maximize the efficiency and productivity of the digitizing operation, (2) utilize the available digitizing budget to the maximum extent (i.e., input the most amount of data for the given budget), and (3) satisfy primary user demand for data. The point of this paper is to proffer the concept that the types of distinct data sets represented by GIS #1 through GIS #5 and their subsets should be viewed as an entire set and ordered for input by a management technique. By applying a number of CFs to each data set (and/or subset) the data base sets or subsets can be ranked. These rankings can be related by some type of decision process (e.g., weighting schemes, decision matrices and so forth) to obtain the final entry priority or order for input to the GIS of the data bases. A discussion of applicable decision methods is beyond the scope of this paper but can be found in standard textbooks.

Each of the following CFs should be applied to each individual data set (or subset) intended for GIS use:

### CF #1. Data Use

Of primary consideration is the ordering or ranking of data sets by relative importance to the primary objective of the GIS and the time and need requirements of the primary user. However, expected use of data set combinations should be also examined. For example, after the user has ranked each data set in order of importance (e.g., geology - #1, vegetation - #2, habitat - #3, wetlands - #4, and so forth), the GIS manager should have combinations of data sets similarly ranked by their most expected use, which is a function of the primary user's intended analyses (e.g., geology/wetlands - #1, vegetation/wetlands - #2). This combination ranking, when compared to the original individual ranking, may significantly influence the data entry priority of the data sets. In the example, wetlands data were not considered a high individual priority data set (#4), but in combination it became important for priority data entry because the primary and secondary analytical analyses of geology/wetlands and vegetation/wetlands could not possibly proceed without the wetlands data set.

### CF #2. Multiple Uses/Users

In general, any data sets that have multiple uses or can be used by multiple users have an intrinsically higher "added" value than data sets restricted to a single use or user. When such is the case, digitizing budgets can often be expanded due to funding participation from multiple user sources.

### CF #3. Unit Digitizing Cost

All data sets should be given a rank in relation to their digitizing cost. Digitizing cost is a direct function of the attributes of the data and mechanics of digitizing the data. The most important data attributes are data denseness and complexity and data reliability and accuracy. Although too elaborate a topic to discuss herein, the most important element pertaining to the mechanics of digitizing is the legibility of data to be digitized (i.e., the overall condition of the source material). Data from good source material can be digitized quicker at a lower unit cost. Relative to data attributes, the least complex data to digitize will have the lowest unit cost and permit the most data entry into the GIS for a given budget. However, the above data and digitizing attributes must be examined in concert with their relationship to unit cost by the GIS manager. If the accuracy of a particular data set is known to be questionable, it is assumed that its intended use will, consequently, be severely restricted or used only with extreme caution. In such a case, regardless of the ease (or low unit cost) of digitizing the data set, the value of the entire data set is in doubt, thereby rendering the cost of the digitizing a possible wasted value that could have been applied to another, more reliable, but, perhaps, more complex data set.

#### CF #4. Data Set Interrelationships

All data sets should be ranked for "stand-alone" usefulness independent of the other data sets. Data sets that are not usable without the availability of other data sets (except perhaps for data sets creating specifically-used visual displays) should be given low rankings. To derive use from such data sets requires the digitization of multiple data sets, thereby increasing unit digitizing time and costs. For example, digitizing manhole cover locations (although of a relatively low cost) can be useless without inputting the entire sewer system map; however, the total cost of digitizing these two data sets may be more appropriately applied to inputting an entire topographic data set as an alternative. Data sets that can be used to infer, imply, or check other data sets by computerized algorithms or interactive user involvement should be given high priority rankings and made more use of to reduce manual editing requirements and streamline quality control of the entire digitization process.

#### CF #5. Data Sales

Data sets should be assessed and ranked according to their potential for sale to sources beyond the user agency. With the advent of wider use of CAD/CAM systems by private and public organizations, data sharing and sales are already experiencing greater demand. Data sets with the highest potential for revenue generation should be given high ranking and priority for digitization. Added revenue can be used to finance further digitization of the other data sets.

#### CF #6. GIS Requirements

The GIS manager must rank data sets considering any limitations imposed by the actual GIS (hardware and software) being utilized and the objective of the GIS (i.e., considering Aangeenbrug's GIS categories #1 through #5). Data entry is often constrained by system limitations. For example, some GIS software does not handle small "islands" extremely well, and in such cases, data sets with numerous "islands" should be given a lower priority for digitizing than other "island-free" data sets. Hardware, particularly scanning versus manual digitizers, may also play an important part in ordering data sets as well.

#### CF #7. Other Data Sources

Occasionally, specific data sets desired for a GIS may be available from a variety of sources but often are at different scales or have other undesirable characteristics. The varying characteristics often compromise the degree of accuracy obtainable



relative to the desired result, and this must be assessed by the GIS manager. An option that is not in widespread use, is the application of computer algorithms to modify existing data sets available from other sources. This modification would be performed in lieu of users digitizing their own original data sets. For example, topographic data tapes available from the Federal Government can be obtained and processed via an interpolation algorithm to refine contour intervals. To a county government, this may be a less expensive alternative than digitizing their own original, larger scale topographic sheets on a countywide basis if the data itself is only intended for use in a Category #3 GIS (i.e., Planning/Evaluation GIS). Data sets are not ranked for this CF, they are simply examined for acquisition from other sources at reduced cost.

The above CFs were not discussed in any order of intended importance nor are actual numerical value rankings (high/low) in each CF suggested. Furthermore, it is acknowledged that there are other CFs or sub-CFs that can be defined and added to the total analysis process. The development of appropriate CFs and entire decision process for management of GIS data base entry is considered to be one that should be individually designed and customized by the GIS manager. The point of this paper, however, was to introduce the application of a systematic management methodology to provide a logical decision process to govern data base entry. The "GIS Data Base Entry Management" concept is a planning tool of major significance that directly influences the amount of data capable of being entered into large-scale GISs within a given budget. With the massive amounts of data to be entered into GISs (which are required to make GISs useful and pay for initial costs), effective data base entry management may become the "Value Engineering" of the entire field.

## SPATIALLY CHARACTERIZING EFFECTIVE TIMBER SUPPLY

J. Berry

Yale University

School of Forestry and Environmental Studies

New Haven, CT 06511

J. Sailor

Intergraph Corporation

Huntsville, AL 35807

### ABSTRACT

The structure of a computer-oriented cartographic model for assessing roundwood supply for generation of base load electricity is discussed. The model provides an analytical procedure for coupling spatial information of harvesting economics and owner willingness to sell stumpage. Supply is characterized in terms of standing timber; of accessibility considering various harvesting and hauling factors; and of availability as affected by ownership and residential patterns. Factors governing accessibility to timber include effective harvesting distance to haul roads as modified by barriers and slopes. Haul distance is expressed in units that take into account the relative ease of travel along various road types to a central processing facility. Areas of accessible timber are grouped into spatial units, termed "timbersheds," of common access to particular haul road segments that belong to unique "transport zones." Timber availability considerations include size of ownership parcels, housing density and excluded areas. The analysis techniques are demonstrated for a cartographic data base in western Massachusetts.

### INTRODUCTION

The demand for wood chips throughout the United States has been increasing in recent years. This increased demand has stimulated new technologies such as whole tree chipping and special chip recovery operations during milling. The significant growth of the pulp and particleboard industries have accounted for most of the increased demand for wood chips. The potential use of chips to generate base load electricity, however, may cause the demand to even more dramatically accelerate over the next decade.

The average price of energy from coal is about fifty percent lower than that from wood. On a national basis it is generally not economical to substitute wood for coal under current economic conditions. On the other hand, some extensively forested regions, such as New England, experience delivered coal and oil prices significantly higher than the national average. In these areas, wood-fired power generation may be economical. In addition, the potential for lower transportation costs and stimulation of local employment, coupled with the lower toxic emissions of wood-fired plants have combined to generate considerable interest in these projects.



The less stringent requirements of species mix for wood-fired power plants and proposed new harvesting techniques necessitate the development of a new methodology for assessing roundwood supply. Most currently used methods treat chip supply as a by-product of milling-oriented operations or as self-sufficient pulping operations on large ownership tracts. Rarely is consideration given to whole tree chipping on diverse ownerships, as will likely be required if wood chips are used in the production of electricity. A complete analysis must integrate consideration of the spatial distribution of timber inventory, harvesting costs, and owners' willingness to sell stumpage.

This paper describes the preliminary structure of a computer-oriented model for assessing effective timber supply. The model spatially characterizes supply in terms of:

- \* inventory of standing timber;
- \* accessibility considering various harvesting and hauling factors; and,
- \* availability as affected by patterns of ownership and development of timber lands.

The emphasis of this study has been on the development of the analytical components of the model, rather than on the calibration of the model. Refinement of model parameters for various physical and economic conditions will be left to subsequent research. Several other papers by the authors have reported on various aspects of this model (Berry and Mansbach, 1981; Berry and Tomlin, 1981; Berry and Sailor, 1981).

#### FUNDAMENTAL CONSIDERATIONS

Information on the extent and location of timber resources serves as primary input. This information can be extracted from several sources, such as existing forest cover type maps, color-infrared aerial photography and LANDSAT satellite imagery. Inventory information alone, however, may not yield accurate estimates of actual roundwood supply as affected by physical considerations. Many forested areas, for example, may be too remote from existing access routes for effective harvesting. Other areas may have highly erodable soils or steep slopes that would likewise make certain harvesting techniques inappropriate. Such areas must be eliminated from consideration or the estimated supply of available roundwood will be too high.

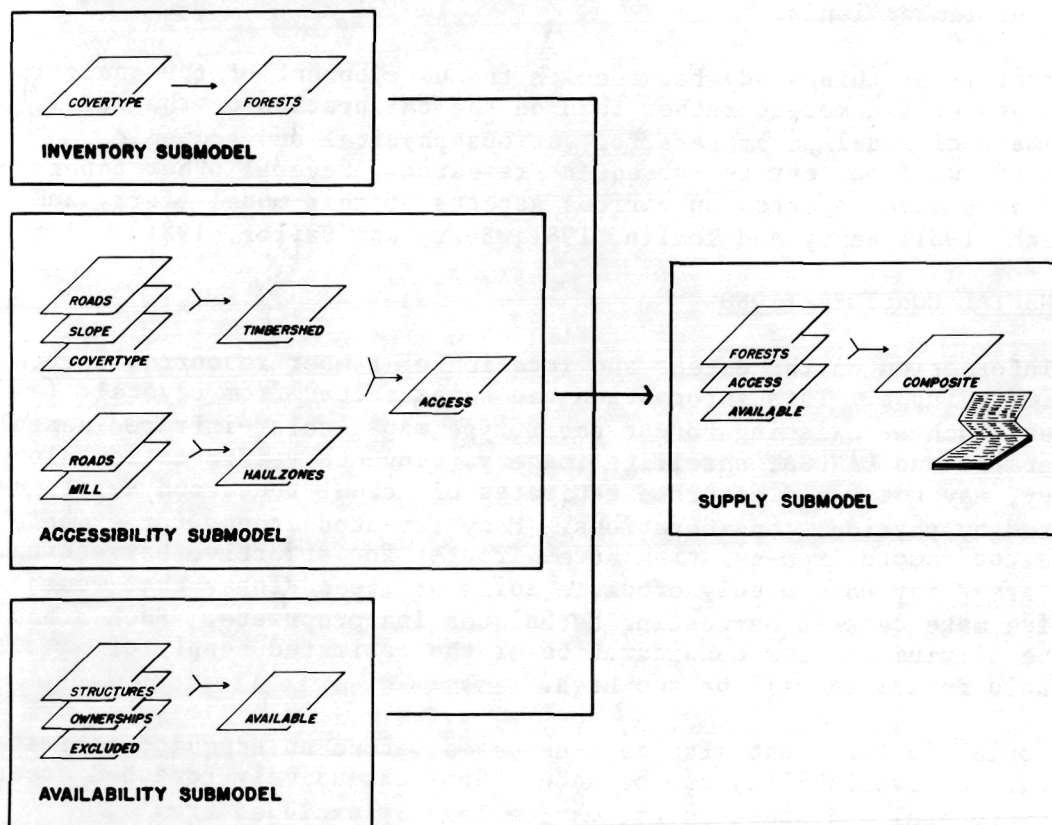
Social factors must also be considered before an adequate estimate of roundwood availability can be made. Some extensively forested areas, such as federal and state parks, may be legally excluded from harvesting. Prudent management practices may also exclude certain areas such as buffer strips around highways and water courses. Of the remaining potentially harvestable areas, ownership characteristics, such as parcel size, can be used to determine propensity of owners to sell stumpage.

The analytic method used to assess timber supply in this manner must be spatially consistent. In addition to producing tables

summarizing supply, maps depicting the distribution of the supply can be invaluable. The method used should be flexible and provide for simulation of various harvesting alternatives and economic environments. It should be easily transported, both in terms of its variables and its processing requirements.

### PRELIMINARY MODEL

This paper summarizes the results of an exploratory study of computer-assisted map analysis techniques to characterize timber supply. Figure 1 is a generalized flow chart of the analytical process. The model consists of four major submodels: inventory, access, availability, and supply. In this preliminary form, only a few considerations for each submodel are included. The "inventory submodel" for this study identifies areas of appropriate forest cover types. The "access submodel" consists of two parts: transport distance, and harvesting characterizations. Transport distance uses a map showing a



**FIGURE 1.** Generalized Flowchart of Model. The cartographic model considers harvesting/hauling access and stumpage availability as well as forest inventory information to spatially characterize the effective timber supply.

proposed mill site and measures haul distance and harvesting accessibility of timber. Haul distance is measured along the existing road network. Constraint maps are used so that distance is measured only along roads and is weighted by road type. For the purpose of this demonstration, timber hauling along secondary roads was assumed to take 50% longer as that along primary roads; hauling on tertiary roads was assumed to take 100% longer than on primary roads; and hauling on unimproved roads was assumed to take 200% longer than on primary roads. From the resulting map of weighted haul distance, a map of haul distance zones was created. A map of accessibility of timber, or effective yarding (skidding) distance to roads in these haul zones, can be generated by taking into consideration steep slopes and water areas that must be circumvented. For the demonstration, slopes were used to weight the distance such that moderately limiting slopes of 11-15% required 100% more time to traverse (and thus for cost purposes are twice as "far" away), severely limiting slopes of 16-20% required 200% more time to traverse, and slopes greater than 20% were avoided. The map resulting from this measurement was converted to show zones of weighted harvesting distance (or cost). The resulting zones are then combined with the vegetation map to produce tables of transport cost versus timber supply.

The second part of the access submodel is concerned with finding minimum yarding distances to haul roads. A map of weighted distance from haul roads is made using the same characteristics as in the transport distance characterization. Using a road intersection map, a map of boundaries of equal haul distance to two or more roads is produced. This map is created by generating lines from road intersections that are constrained to being on the weighted distance margin between two roads. These boundaries are areas that are equally accessible to two or more roads, using the weighted accessibility measure. On either side of these boundaries, timber is more accessible to a particular haul road, and therefore will cost less to harvest if it is yarded to that road.

The "availability submodel" considers housing density in characterizing the propensity of owners to sell stumpage. Areas with a relatively high density are ranked as having a lower likelihood of being available for harvesting. Those areas of relatively low housing density, conversely, are considered more likely to be sold for stumpage. Maps of housing density are derived from maps identifying residential and commercial structures within a study area. For purposes of this study, the total number of structures with a radius of 1/8 mile was used to indicate the relative housing density for an area. A second consideration of availability involves ownership parcel size. If an area falls within a large parcel, that is, greater than 10 hectares, it is considered more likely to be available. The submodel also eliminates from consideration any areas which prohibit harvesting.

The maps of harvesting potential and owners' propensity to sell are combined in the "supply submodel" to characterize the forest resource in terms of effective supply. The primary output of this submodel are



tables summarizing the areal extent of each forest type in each accessibility/availability class. In addition, maps locating the various combinations can be generated. Contiguous units of user-defined classes are identified and their areal extent computed. Parcels that are too small for efficient harvesting can be easily identified and eliminated.

#### SPATIAL DATA BASE

Information used in this study is part of a general purpose data base being developed for the Harvard Forest vicinity (Petersham, Massachusetts). Each map represents 1770 hectares or 17.7 square kilometers. (1/4 hectare per cell, 7080 cells in all). The maps used for this study include vegetation cover types, elevation, roads, and water features.

Data encoding, analysis, and display capabilities for this study were provided through the use of software developed at Yale University as part of the Map Analysis Package (Tomlin, in preparation). Information on the biological, physical, and cultural features of a given geographic area is encoded to correspond with a grid cell data structure. Each grid cell is assigned a value which represents one member of a set of mutually exclusive categories (e.g. dry land, stream, pond, lake). These data are analyzed using a flexible package of fundamental processing operations that are logically sequenced to form a cartographic model (Tomlin and Berry, 1979; Berry and Tomlin, 1982).

#### DEMONSTRATION RESULTS

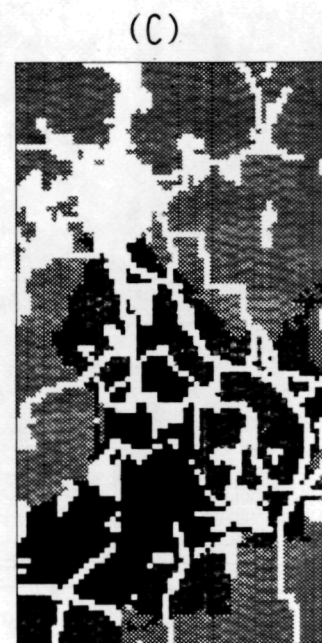
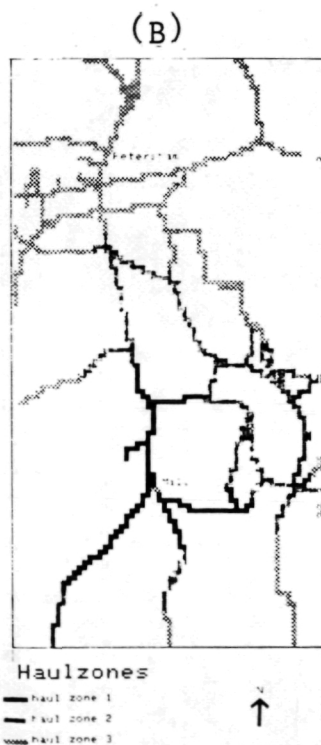
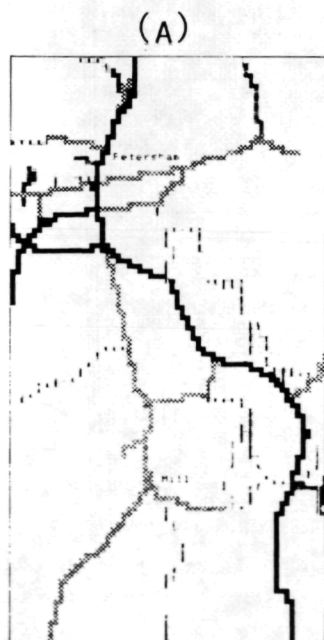
The thirty-four vegetation types occurring in the Petersham area were collapsed into nine classes of merchantable forests. For display purposes, these are grouped into five categories (Figure 2). Forested areas comprise 83.6% of the study area. However, these areas have different accessibility and transport costs which must be considered in determining potential supply.

Figure 3 shows important intermediate maps associated with the transport submodel. For display and tabulation the road network was divided into three zones of haul distance, and the timber areas were divided into two zones of accessibility. Because the distances measured are a function of the roads or terrain traversed, the maps can be considered travel time maps or transport cost surfaces. The units can be expressed in time, cost, or in distance equivalents. In this case, "distances" are expressed relative to kilometers travelled on primary roads for haul distance and kilometers skidded on level ground for accessibility. Figure 4 shows intermediate maps from the "timbersheds" submodel. The results of the access submodel allow the strategic planner to characterize the timber supply surrounding a mill site in terms of harvesting costs. This information can then be used to determine if the supply, within economic transport and access "reach", is sufficient to sustain the proposed facility. Table 1 shows the forest types as a function of harvesting and hauling considerations.



**Figure 2.** Forest Type Map. The nine forest classes for the study are grouped into five categories for display.

FORESTS	NON-FORESTS	CELLS	COVERAGE
1	2	1281 CELLS	16.42 COVERAGE
3	4	1281 CELLS	1.31 COVERAGE
5	6	91 CELLS	1.35 COVERAGE
7	8	1281 CELLS	2.75 COVERAGE
9	10	1281 CELLS	4.32 COVERAGE
11	12	282 CELLS	4.35 COVERAGE
13	14	943 CELLS	13.35 COVERAGE
15	16	1071 CELLS	26.45 COVERAGE
17	18	1909 CELLS	27.45 COVERAGE
19	20, 21, 22, 23	213 CELLS	3.32 COVERAGE



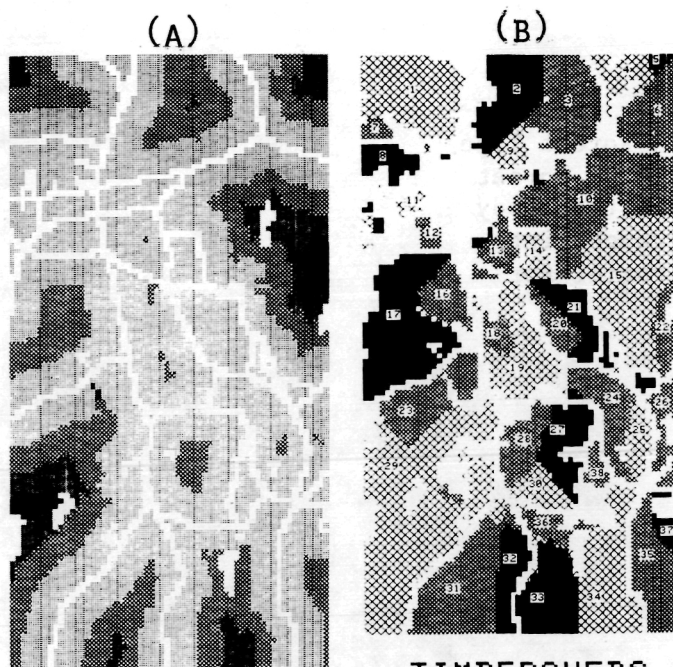
Transport Zones

- closest to zone 1
- closest to zone 2
- closest to zone 3
- non-forest



1 km

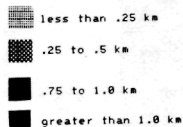
**Figure 3.** Transport Analysis Maps. The existing roads (a), weighted haul distance along these roads (b), and the final transport zones, (c) are shown.



**Figure 4. Harvesting Analysis Maps.** Accessibility of timber characterized as weighted skidding proximity (a) and areas of common access to haul road segments (b) are shown.

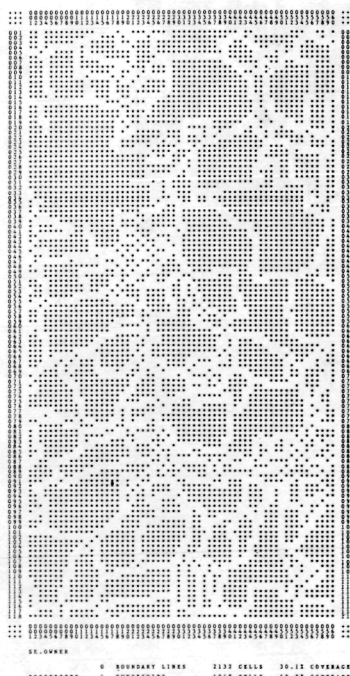
#### HARVEST ACCESS

EFFECTIVE SKIDDING DISTANCE



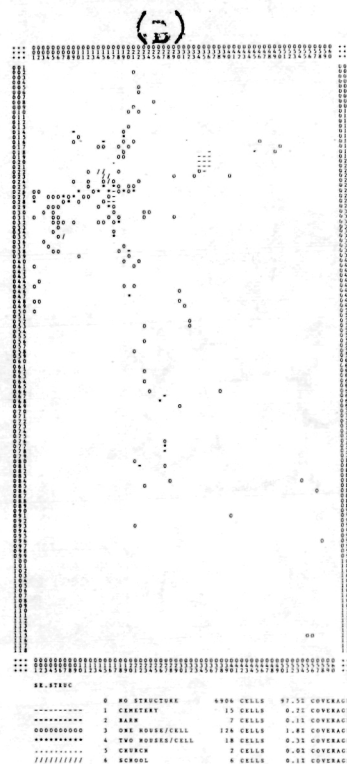
N  
↑

1 km



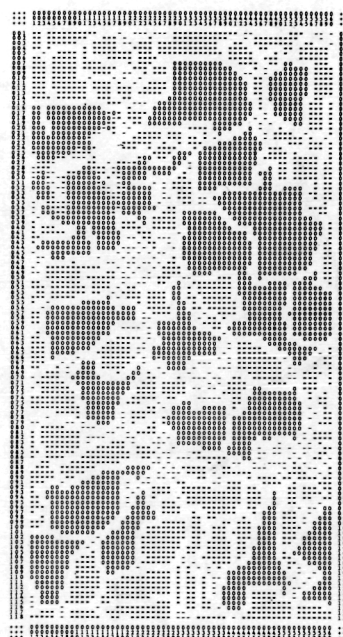
SE. OWNER  
0 BOUNDARY LINES 2122 CELLS 30.12 COVERAGE  
4 OVERSHOOTS 4248 CELLS 40.92 COVERAGE

#### TIMBERSHEDS



NO. STRUCT  
0 NO STRUCTURE 4204 CELLS 97.55 COVERAGE  
1 CEMETARY 15 CELLS 0.32 COVERAGE  
2 BARN 7 CELLS 0.15 COVERAGE  
3 ONE HOUSE/CELL 114 CELLS 1.81 COVERAGE  
4 TWO HOUSES/CELL 18 CELLS 0.32 COVERAGE  
5 CHURCH 2 CELLS 0.01 COVERAGE  
6 SCHOOL 6 CELLS 0.12 COVERAGE

#### (C)



LIBERTY  
0 NOT APPLIC. 3194 CELLS 31.85 COVERAGE  
1 LOW APPLIC. 375 CELLS 5.35 COVERAGE  
2 MOD. APPLIC. 2822 CELLS 39.42 COVERAGE  
3 HIGH APPLIC. 2491 CELLS 33.35 COVERAGE

**Figure 5. Availability Analysis Maps.** Maps of ownership patterns (a) and residential dwellings (b) are used to characterize the likelihood of areas to be sold for stumpage. Combining these three maps yields a map (c) of the relative availability for harvesting.

(A)



Table 1 -- Tabulation of forest classes by transport distance zones

Forest <sup>2</sup> Class	Total Area(ha)	Transport Zone <sup>1</sup> (Haul/Access) (ha)					
		1/1	1/2	2/1	2/2	3/1	3/2
1	26.25	0	0	3.5	0	21.75	1.25
2	22.75	1.75	0	10.5	0	9.25	1.25
3	47.0	2.0	0	10.75	0.75	11.5	22.0
4	76.0	7.75	0	12.25	0	44.25	11.75
5	73.0	11.0	0.75	23.75	0	21.25	16.25
6	235.75	21.25	1.5	48.5	0	132.0	32.5
7	477.25	45.5	5.25	142.0	1.5	194.75	78.5
8	53.75	104.0	4.25	83.0	3.75	177.75	52.0
9	53.75	9.0	0.5	14.0	2.0	12.75	17.0

<sup>1</sup>Transport Zones (for example, 1/2 is Haul Zone 1, Access Zone 2)

#### Haul Zones

Haul Zone 1 - less than 2 km. haul

Haul Zone 2 - 2 km. to 4 km. haul

Haul Zone 3 - greater than 4 km. haul

Haul distances weighted by road type and expressed relative to hauling on primary road.

#### Access Zones

Access Zone 1 - less than .5 km skid

Access Zone 2 - .5 km skid or greater

Skid distance weighted by slope and expressed relative to skidding on flat surface

#### <sup>2</sup> Forest Classes

Class 1 - hardwoods; 41-60 ft.; 81-100% closure

Class 2 - hardwoods; 61-80 ft.; 81-100% closure

Class 3 - softwoods; 21-60 ft.; 30-80% closure

Class 4 - softwoods; 61-80 ft.; 81-100% closure

Class 5 - mixedwoods; (S/H); 21-60 ft.; 30-80% closure

Class 6 - mixedwoods; (S/H); 61-80 ft.; 30-80% closure

Class 7 - mixedwoods; (H/S); 21-80 ft.; 30-80% closure

Class 8 - mixedwoods; (H/S); 61-80 ft.; 30-80% closure

Class 9 - mixedwoods; (uneven aged)

Figure 5 shows the maps of the availability submodel. Ownership parcels of more than 20 hectares comprise 37.4% of the study area. These larger parcels can be considered as having a high probability of being sold for stumpage. Similarly areas of less than five structures per 0.5 square kilometer (approximately one house per 10 hectares) are considered as being likely for sale. These relatively unpopulated areas comprise 97.0% of the study area. Combining these two maps creates a map of overall availability. This map identifies 35.2% of the study area as being likely to be available for stumpage. However, some of these areas may actually be unavailable due to legal statute or management policy. For this demonstration, areas to be excluded from harvesting include institutional areas and park lands. These comprise only 9% and are spatially coincident with populated areas in most instances. As a result the consideration of excluded areas only slightly decreases the "likely to be available" areas to 35.1%.

The final phase of the model combines the information on access and availability for the merchantable forested areas. Figure 6 locates the forested areas that have good access and are likely to be available. Of the 1480 hectares of merchantable forests only 200 hectares are in this desirable category. In addition, with the exception of a few tracts, most of these areas are well dispersed and relatively small. Maps similar to the one in Figure 6 can be displayed for any of the various combinations of accessibility and availability of the forested areas. The total amount of forested areas which meets the minimum requirements of this analysis is 960 hectares. The purely physical inventory of timberlands greatly overstated this acreage and offered no information as to the relative desirability or spatial distribution of the remaining land.

## CONCLUSION

An advantage of computer-assisted map analysis is that once a model is developed and the appropriate data encoded, repeated simulation of the model using different calibration coefficients yields insight into the unique character of an area. For example, if effective skidding distance is extended from 0.3 kilometers to 0.5 kilometers and parcel size reduced from 20 hectares to 10 hectares, the highly desirable forested acreage increases from 200 hectares to 325 hectares. This method of sensitivity analysis can be used to identify the more important considerations as well as give a range of expected supply under various engineering and economic environments.

The model serves as an excellent strategic planning tool. It locates general areas of likely accessible and available forests and provides insight into the significant factors affecting potential supply. The analysis, however, is not intended to provide output useful to the harvesting crew. Rather it is intended to better indicate actual timber supply than conventional inventory-driven procedures.





**Figure 6. Effective Timber Supply.** This map depicts the forested areas that are likely to be available and easily accessible by forest cover classes 1-9 (see Table 1). Although 84% of the study area has forest cover, this analysis shows that only 11% is in the most desirable category.

AVANCE			NOTA CELLS	PERCENT COVERAGE
0	PPF-0,PPF-0,PPF-0	29 CELLS	0.45 COVERAGE	
1	PPF-1,PPF-1,PPF-1	1 CELLS	0.01 COVERAGE	
2	PPF-2,PPF-2,PPF-2	7 CELLS	0.11 COVERAGE	
3	PPF-3,PPF-3,PPF-3	13 CELLS	0.20 COVERAGE	
4	PPF-4,PPF-4,PPF-4	41 CELLS	0.62 COVERAGE	
5	PPF-5,PPF-5,PPF-5	123 CELLS	1.85 COVERAGE	
6	PPF-6,PPF-6,PPF-6	348 CELLS	5.35 COVERAGE	
7	PPF-7,PPF-7,PPF-7	148 CELLS	2.25 COVERAGE	
8	PPF-8,PPF-8,PPF-8	27 CELLS	0.41 COVERAGE	

## ACKNOWLEDGEMENTS

The research described in this paper was supported by grants from the General Services Foundation and internal Yale University developmental research funds.

## REFERENCES

- Berry, J.K. and A. Mansbach. 1981. Extending the Utility of Forest Cover Maps. Proceedings of The Eastern Regional Remote Sensing Applications Conference, Danvers, Massachusetts.
- Berry, J.K. and C.D. Tomlin. 1981. A Cartographic Model for Assessing Roundwood Availability. In: WOODPOWER: New Perspective in Forest Usage, Pergamon Press.
- Berry, J.K. and J.K. Sailor. 1981. A Spatial Analysis of Timber Supply. Proceedings of The In-Place Resource Inventories: Principles and Practices, University of Maine, Orono, Maine. (In press).
- Berry, J.K. and C.D. Tomlin. 1982. Cartographic-Assisted Analysis of Spatially Defined Neighborhoods. Proceedings of The National Conference on Energy Resource Management, Amer. Planning Assoc., Vienna, Virginia.

Tomlin, C.D. (in preparation). Digital Cartographic Modeling Techniques in Environmental Planning. Doctoral Dissertation, Yale School of Forestry, New Haven, Connecticut.

Tomlin, C.D. and J.K. Berry. 1979. A Mathematical Structure for Cartographic Modeling in Environmental Analysis. Proceedings of the American Congress on Surveying and Mapping, 39th Meeting, Falls Church, Virginia.

CARTOGRAPHIC MODELING: COMPUTER-ASSISTED ANALYSIS OF SPATIALLY  
DEFINED NEIGHBORHOODS

J.K. Berry  
Yale University  
School of Forestry and Environmental Studies  
205 Prospect Street  
New Haven, CT 06511

C.D. Tomlin  
Harvard University  
Graduate School of Design  
Cambridge, MA 02138

ABSTRACT

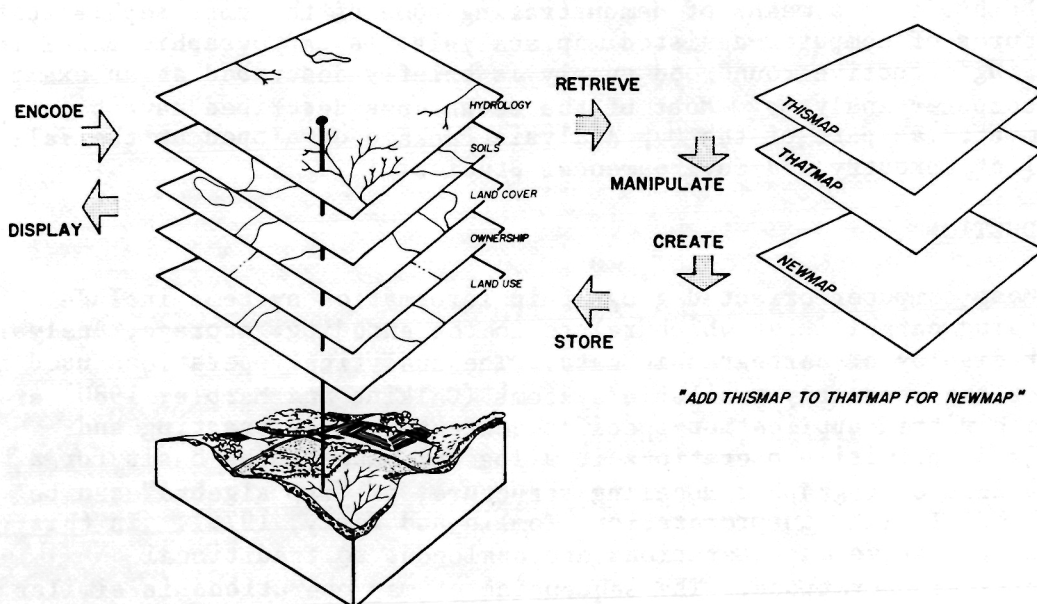
Cartographic models addressing a wide variety of applications are composed of fundamental map processing operations. These primitive operations are neither data base nor application-specific. By organizing the set of operations into a mathematical-like structure, the basis for a generalized cartographic modeling framework can be developed. Among the major classes of primitive operations are those associated with reclassifying map categories, overlaying maps, determining distance and connectivity, and characterizing cartographic neighborhoods. This paper establishes the conceptual framework of cartographic modeling and uses techniques for characterizing neighborhoods as a means of demonstrating some of the more sophisticated procedures of computer-assisted map analysis. A cartographic model for assessing effective roundwood supply is briefly described as an example of a computer analysis. Most of the techniques described have been implemented as part of the Map Analysis Package developed at the Yale School of Forestry and Environmental Studies.

INTRODUCTION

Most computer-oriented geographic information systems include processing capabilities which relate to the encoding, storage, analysis and/or display of cartographic data. The analytical operations used in many of the currently available systems (Calkins and Marble, 1980) are embedded within application-specific contexts. By extracting and organizing primitive operations in a logical manner, the basis for a generalized cartographic modeling structure, or "map algebra" can be developed (Tomlin, in preparation; Tomlin and Berry, 1979). In this context primitive map operations are analogous to traditional mathematical operations. The sequencing of map operations is similar to the algebraic solution of equations to find unknowns. In this case, however, the unknowns represent entire maps. The conceptual framework interrelating these primitive operations provides a basis for a modeling structure which accommodates a wide variety of computer analyses. This paper describes this conceptual framework and uses the techniques for characterizing cartographic neighborhoods to demonstrate some of the more sophisticated procedures and considerations of cartographic modeling.

## DATA AND PROCESSING STRUCTURES

In order to use primitive operations in a modeling context a common data structure and a flexible processing structure must be used. The variety of mappable characteristics likely to be associated with any given geographic location may be organized as a series of spatially registered computer-compatible maps (Figure 1). In this way, a data base may be defined as a set of maps registered over a common geographic area; a map, or "overlay", may be defined as a set of mutually exclusive but thematically related categories; and a category, or "region", may be defined as a thematic value associated with a set of geographic locations, or "points" (Tomlin and Tomlin, 1981). While this is certainly not the only way to represent cartographic data (Chrisman and Peucher, 1975) it is one which relates directly and intuitively to traditional graphic techniques involving conventional geographic maps. It is also one which is common to many computer-oriented geographic information systems. Differences among these systems relate to either the way in which thematic attributes are represented (i.e. numerically, literally or in binary form) or to the way in which locational attributes are coded (i.e. rectangular cells, polygons, line segments, etc.). While these differences are significant in terms of implementation strategies, they need not affect the definition of fundamental cartographic techniques.



**Figure 1.** Cartographic Modeling Concept. A data base consists of spatially registered maps. Cyclical processing of these data involves retrieving one or more maps which are used to create a new map. The derived map then becomes available for subsequent processing.



If primitive operations are to be flexibly combined, each must accept input and generate output in the same format. Using a data structure as outlined above this may be accomplished by requiring that each analytic operation involve:

- \* retrieval of one or more maps from the data file;
- \* manipulation of that data;
- \* creation of a new map whose categories are represented by thematic values defined as a result of that manipulation; and,
- \* storage of that new map for subsequent processing.

The cyclical nature of this processing structure (Figure 1) is analogous to the evaluation of "nested parentheses" in traditional algebra. The logical sequencing of primitive operations on a set of maps forms a cartographic model of a specified application. As with traditional algebra, fundamental techniques involving several primitive operations can be identified (e.g. a "travel-time" map) that are applicable to numerous situations. The use of primitive analytical operations in a generalized modeling context accommodates a variety of analyses in a common, flexible and intuitive manner. It also provides a framework for instruction in the principles of computer-assisted map analysis that stimulates the development of new techniques and applications (Berry and Tomlin, 1980).

#### FUNDAMENTAL OPERATIONS

Within the data and processing structures outlined above, each primitive operation may be regarded as an independent tool limited only by the general thematic and/or spatial characteristics of the data to which it is applied. From this point of view; four major classes of fundamental map analysis operations may be identified (Table 1). These involve:

- \* reclassifying map categories;
- \* overlaying maps;
- \* determining distance and connectivity; and,
- \* characterizing cartographic neighborhoods.

A brief discussion of these fundamental classes is presented below. More detailed discussions are presented in several of the references noted at the end of this paper (Berry and Tomlin, 1982; Berry, 1981; Tomlin and Berry, 1979).

The first of the four major groups of cartographic modeling operations is the simplest and, in many ways, the most fundamental.

**TABLE 1 FUNDAMENTAL MAP ANALYSIS OPERATIONS**

FUNDAMENTAL CLASSES	FUNCTIONAL BASIS	EXAMPLE OPERATIONS
RECLASSIFYING MAP CATEGORIES- operations for reclassifying map categories involve reassigning thematic values to the categories of an existing map as a function of the initial value, the position, the size or the shape of the spatial configuration associated with each category.	<ul style="list-style-type: none"> <li>• INITIAL VALUE</li> <li>• POSITION</li> <li>• SIZE</li> <li>• SHAPE</li> </ul>	<ul style="list-style-type: none"> <li>• ARBITRARY SCHEME (relabeling, isolating, aggregating)</li> <li>• ORDERING SCHEME (ranking, weighting)</li> <li>• MATHEMATICAL RULE (isolating, arithmetics with constants)</li> <li>• SPATIAL LOCATION (reference coordinates, line orientation)</li> <li>• AREAL EXTENT</li> <li>• VOLUME</li> <li>• BOUNDARY CONFIGURATION (edginess, irregularity)</li> <li>• SPATIAL INTEGRITY (interior holes, fragmentation)</li> </ul>
OVERLAYING MAPS-- overlay operations result in the creation of a new map where the values assigned to every location on that map is computed as a function of independent values associated with that location on two or more existing maps.	<ul style="list-style-type: none"> <li>• LOCATION-SPECIFIC</li> <li>• CATEGORY-WIDE</li> <li>• MAP-WIDE</li> </ul>	<ul style="list-style-type: none"> <li>• PERMUTATION (category combination)</li> <li>• DIVERSITY COUNTING</li> <li>• RELATIVE PROPORTION (frequency)</li> <li>• ORDINAL SELECTION (maximize, minimize, median, etc)</li> <li>• MASKING/SIEVING</li> <li>• LOGICAL COMBINATION (union, intersection)</li> <li>• ARITHMETIC COMBINATION (add, subtract, divide, etc)</li> <li>• WEIGHT AVERAGING</li> <li>• PERMUTATION (category combination)</li> <li>• DIVERSITY COUNTING</li> <li>• RELATIVE PROPORTION (frequency, uniqueness, overlap)</li> <li>• ORDINAL SELECTION (maximize, minimize, median, etc)</li> <li>• LOGICAL COMBINATION (union, intersection)</li> <li>• ARITHMETIC COMBINATION (add, multiply, etc- commutative only)</li> <li>• SOLUTION OF MATHEMATICAL/STATISTICAL RELATIONSHIPS</li> </ul>
DETERMINING DISTANCE AND CONNECTIVITY- operations for measuring cartographic distance involve the creation of new maps in which the distance and route between points can be expressed as simple Euclidean length or as a function of absolute and/or relative barriers.	<ul style="list-style-type: none"> <li>• SIMPLE DISTANCE/PROXIMITY</li> <li>• WEIGHTED PROXIMITY</li> <li>• CONNECTIVITY</li> </ul>	<ul style="list-style-type: none"> <li>• SHORTEST STRAIGHT LINE ("as-the-crow-flies")</li> <li>• SHORTEST ROUTE ("as-the-crow-walks")</li> <li>• ABSOLUTE BARRIERS (guide movement)</li> <li>• RELATIVE BARRIERS (expend units of movement)</li> <li>• STRAIGHT LINE (simple distance, intervisibility)</li> <li>• OPTIMAL PATH (steepest downhill path, minimal "cost" route)</li> <li>• OPTIMAL PATH DENSITY (networks)</li> </ul>
CHARACTERIZING CARTOGRAPHIC NEIGHBORHOODS- these operations involve the creation of a new map based on the consideration of "roving windows" of neighboring points about selected target locations.	<ul style="list-style-type: none"> <li>• SUMMARIZING THEMATIC ATTRIBUTES</li> <li>• 2-DIMENSIONAL FEATURE ATTRIBUTES</li> <li>• 3-DIMENSIONAL SURFACE ATTRIBUTES</li> </ul>	<ul style="list-style-type: none"> <li>• STATISTICS (total, mean, maximum, etc)</li> <li>• ANOMALY DETECTION (deviation, proportion similar)</li> <li>• INTERPOLATION (weighted average, nearest neighbor)</li> <li>• MAP GENERALIZATION (surface fitting)</li> <li>• NARROWNESS (shortest cord to opposing edges)</li> <li>• CONTIGUITY (individual clumps)</li> <li>• SLOPE (topographic slope, differentiation)</li> <li>• ORIENTATION (topographic aspect, direction of movement)</li> <li>• PROFILE (patterns along sequential cross-sections)</li> </ul>

Each of the operations involves the creation of a new map by reassigning thematic values to the categories of an existing map. These values may be assigned as a function of the initial value, the position, the size, or the shape of the spatial configuration associated with each category. All of the reclassification operations involve the simple "repackaging" of information on a single map and results in no new boundary delineations.

Operations for overlaying maps begin to relate to the spatial, as well as to the thematic nature of cartographic information. Included in this class of operations are those which involve the creation of a new map such that the value assigned to every location is a function of the independent values associated with that location on two or more existing maps. In simple location-specific overlaying, the value assigned is a function of the spatially aligned coincidence of the existing maps. In category-wide compositing values are assigned to entire thematic regions as a function of the values associated with the regions contained on the existing maps. Whereas the first overlaying approach conceptually involves "vertical spearing" of a set of maps, the latter approach uses one map to identify boundaries from which information is extracted in a "horizontal summary" fashion from the other maps. A third overlay approach treats each map as a variable; each location as a case and each value as an observation in evaluating a mathematical or statistical relationship.

The third class of operations is one which relates primarily to the locational nature of cartographic information. Operations in this group generally involve the measurement of distance and the identification of routes between locations on a map surface. This class of operations served as the focus of a recent paper by the authors (Berry and Tomlin, 1982).

The simplest of these operations involves the creation of a map in which the value assigned to each location indicates the shortest distance "as the crow flies" between that location and a specified target area. The result is a map of concentric, equidistant zones around the target area. The target area is not constrained to a single location and can be comprised of a set of dispersed points, lines or areal features.

If movement is implied in the measurement of distance the shortest route between two points may not always be a straight line. And even if it is straight, the Euclidean length of that line may not always reflect a meaningful measure. Rather, distance may be defined in terms of factors such as travel-time, cost, or energy which, unlike miles, may be consumed at rates which vary over space and time. Distance-modifying effects may be expressed cartographically as absolute and relative "barriers" located within the space over which distance is being measured. The resultant map identifies an effective proximity surface that characterizes movement from a target area over that space and through those barriers.

A distance-related set of operations determines the connectivity among specified locations. One such operation traces the steepest downhill path from a point on a three-dimensional surface. For a topographic surface, the path would indicate surficial water flow. For a surface represented by a travel-time map, this can be used to trace minimum-time (i.e. quickest) path. Another operation determines connectivity measured only for straight rays emanating from a target area over a three-dimensional surface to identify visual exposure.

The fourth and final group of operations includes procedures that create a new map in which the value assigned to a location is computed as a function of the independent values within a specified distance around that location (i.e., its neighborhood). This class of techniques will be discussed in detail in the remaining sections of this paper.

#### CHARACTERIZING CARTOGRAPHIC NEIGHBORHOODS

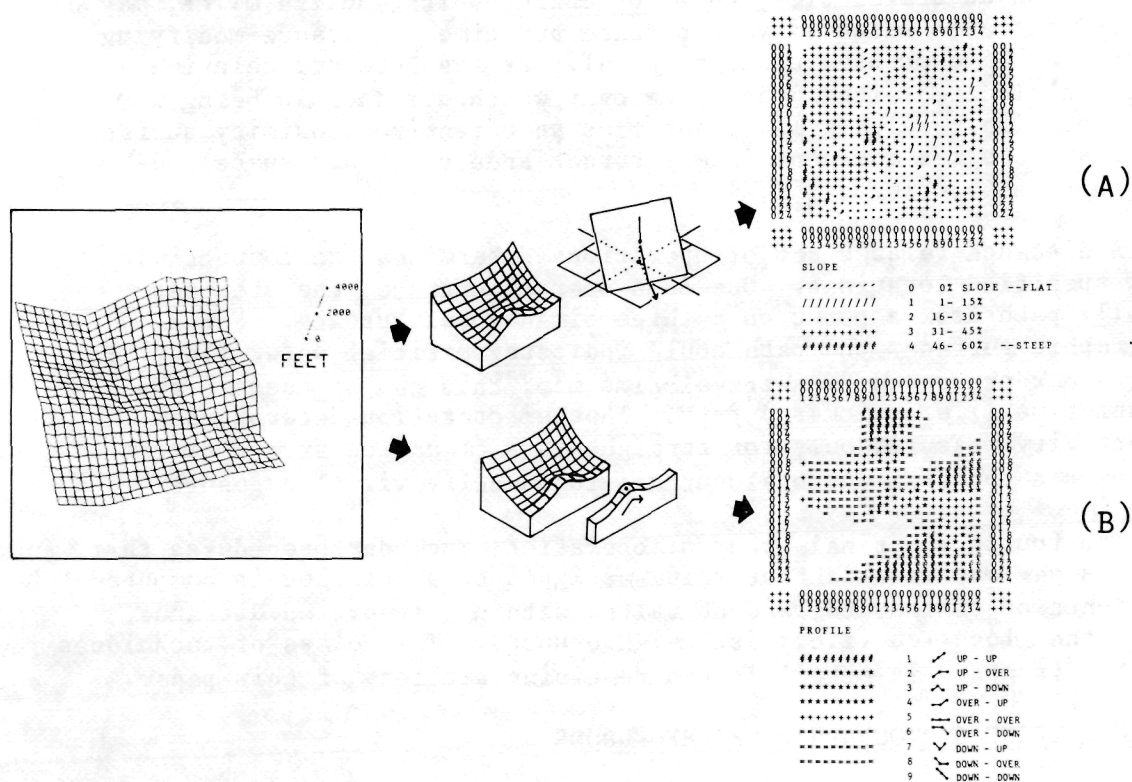
Most geographic information systems contain analytic capabilities for reclassifying and overlaying maps. These operations address the majority of applications that parallel conventional map analysis techniques (McHarg, 1969). However, to more fully integrate spatial considerations with contemporary analysis and planning, new techniques are emerging. The consideration of a location in context with its neighboring locations identifies a set of advanced operations. The summary of information within the neighboring locations can be based on



the configuration of the surface (e.g. slope and aspect), the characterization of contiguous features (e.g. narrowness) or the statistical summary of thematic values (e.g. average value).

The initial step in characterizing cartographic neighborhoods is the establishment of neighborhood membership. A neighborhood, or "roving window," is uniquely defined for each target point as the set of all points which lie within a specified distance and direction around it. In most applications the window has a uniform geometric shape and orientation (e.g. a circle or square). However, as noted above that distance may not necessarily be Euclidean nor symmetrical, such as a neighborhood of "down-wind" locations from a smelting plant.

The characterization of a neighborhood may be based on the relative spatial configuration of values that occur with the neighborhood. This is true of operations which measure topographic characteristics, such as slope, aspect or profile from elevation values. A frequently used techniques involves the "least squares fit" of a plane to adjacent



**Figure 2.** Characterizing Surface Configuration. Least squares fit of a plane to elevation values determines slope and aspect (a); computed cross-sectional profile as viewed toward the northeast (b).



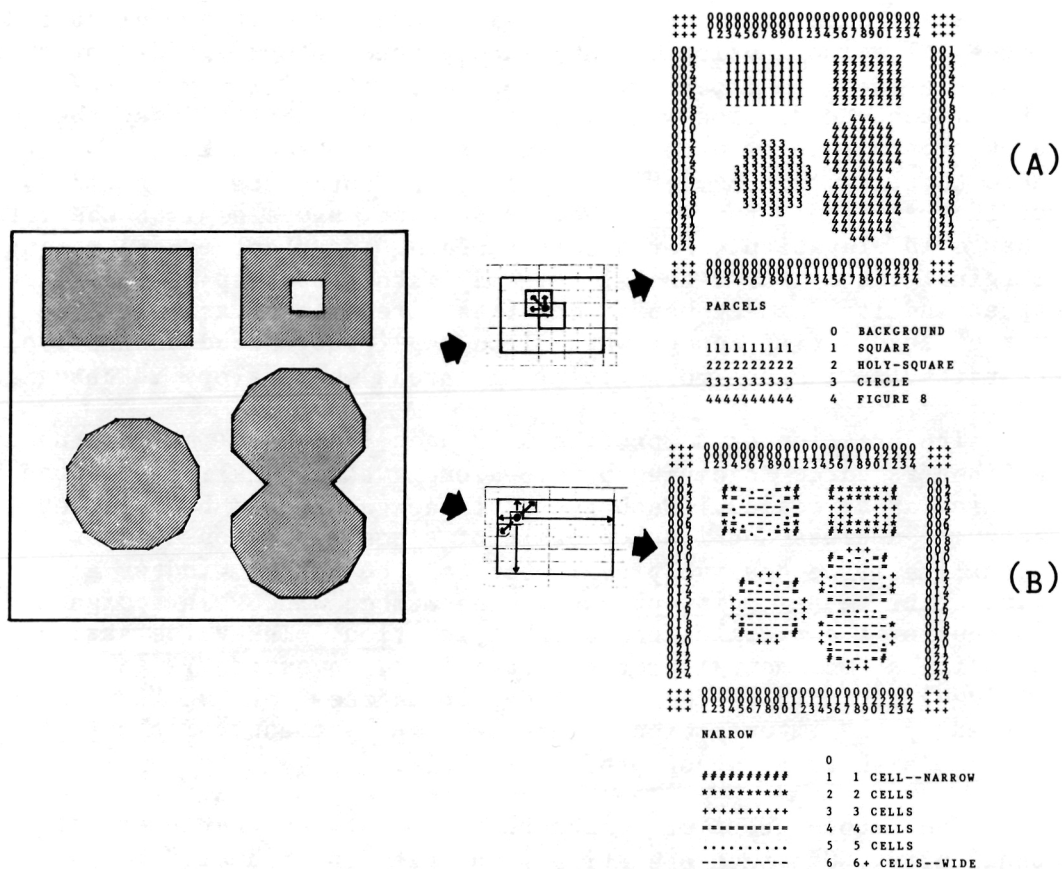
elevation values. This process is similar to fitting a linear regression line to a series of points expressed in two-dimensional space. The inclination of the plane denotes slope and its orientation characterizes the aspect within the immediate vicinity of the focus of the neighborhood. The window is successively shifted over the entire elevation map to produce a continuous slope (Figure 2a) or aspect map. Note that the "slope map" of any surface represents the first derivative of that surface. For an elevation surface, slope depicts the rate of change in elevation. For a cost surface, its slope map represents marginal cost. For a travel time map, its slope map indicates relative speed and its "aspect map" identifies direction of travel. The slope map of an existing topographic slope map (e.g. second derivative) will characterize surface roughness (e.g. areas where slope is changing).

The creation of a "profile map" uses individual neighborhoods defined as three adjoining points along a straight line oriented in a particular direction. Each set of three values can be regarded as defining a cross-sectional profile of a small portion of that surface. Each line is successively evaluated for the set of windows along the line. The center point of each three member window is assigned a value indicating the profile form at that location. The value assigned can identify a fundamental profile class (e.g. inverted "v" shape indicating a ridge) or indicate the magnitude, in degrees, of the "skyward angle" formed by the intersection of the two line segments of the profile. Figure 2b shows a map of profile changes.

The second group of neighborhood operations characterizes contiguity. One such operation identifies individual "clumps" of one or more points that are geographically connected. This involves noting the association between a "target" point and each point of similar thematic value which lies within its neighborhood. If this is done for all points of a given value on a map, spatially contiguous or near-contiguous subsets of those points can be identified. For example, given a map of many lakes this might be used to uniquely identify a particular lake.

The processing technique defines a window that includes neighboring points above and to the left of a location (Figure 3a). This window is successively moved from left to right beginning at the top of the map and proceeding to the bottom. The value assigned is determined according to the sequence in which individual clumps are encountered. If the thematic value of a target point is the same as a member of its neighborhood, it will be assigned the same clump number. If it is not the same, a new clump is indicated. The procedure assigns a common clump number to any groupings that are found to join in lower a portion of the map.

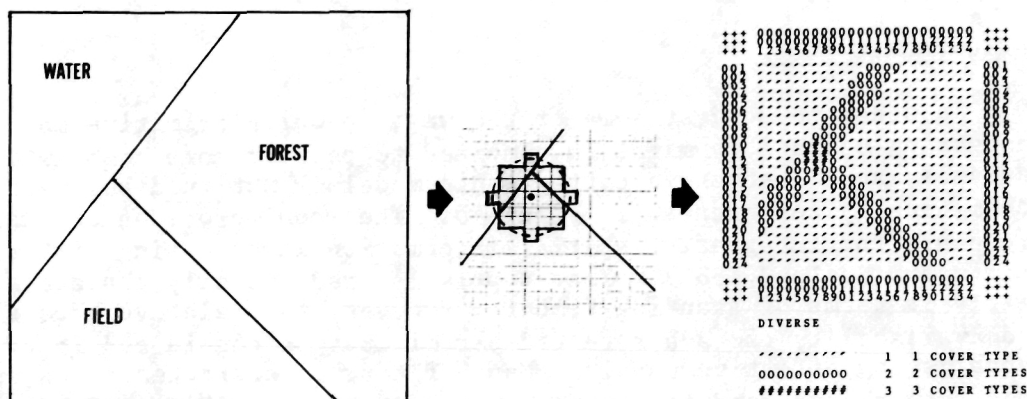
Another neighborhood characteristic which relates to spatial contiguity is narrowness. The narrowness at each point within a map feature is defined as the length of the shortest line segment which can be constructed through that point to diametrically opposing edges of the



**Figure 3.** Characterizing Feature Contiguity. Individual parcels of a common theme can be identified (a); narrowness of features is computed as the shortest cord through each point connecting the border (b).

feature. The state of Massachusetts, for example, is generally narrowest in the vicinity of Cape Cod. In establishing narrowness a window is defined in which the distance from a target location to each feature boundary location is computed (Figure 3b). The total length of each cord passing through the target point is the sum of the distance from that point to opposing boundary locations. The shortest of these cords identifies narrowness at that location. In order to avoid unnecessary processing for some applications, a window of maximum narrowness to be considered is specified.

The final class of neighborhood operations are those that summarize thematic values. Among the simplest of these involve the calculation of summary statistics associated with the map categories occurring within each neighborhood. These statistics might include, for example, the maximum income level, the minimum land value, the diversity of



**Figure 4.** Summarizing Thematic Values. The diversity of cover types within a specified distance can be computed. During processing a "roving window" is used to establish the set of neighboring points used in the summary.

vegetation within a half-mile radius, or perhaps a five-minute radius, of each target point (Figure 4). They might also include the total, the average, or the median value occurring within each neighborhood; the standard deviation or variance of those values; or the difference between the value occurring at a target point itself and the average of those around it.

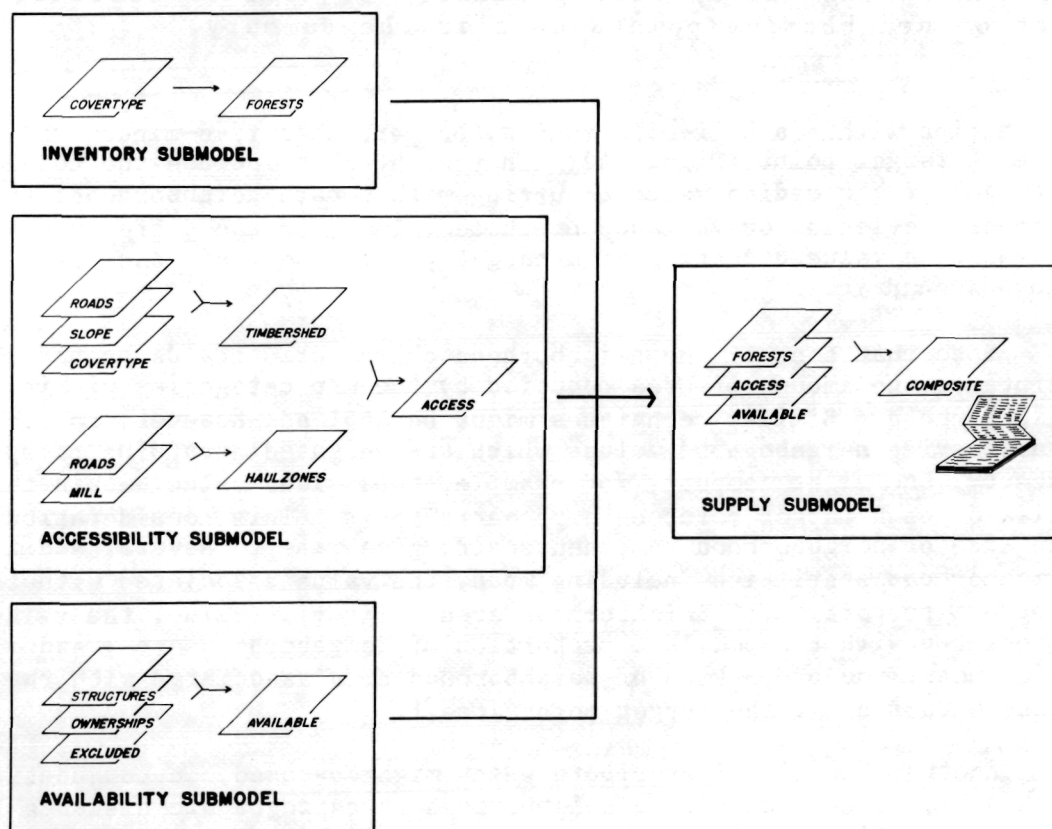
Note that none of the neighborhood characteristics described so far relate to the amount of area occupied by the map categories within each neighborhood. Similar techniques might be applied, however, to characterize neighborhood values which are weighted according to spatial extent. One might compute, for example, total land value within three miles of each target point on a per-acre basis. This consideration of the size of neighborhood components also give rise to several additional neighborhood statistics including mode, the value associated with the greatest proportion of neighborhood areas; minority value, the value associated with the smallest proportion of neighborhood area; and uniqueness, the proportion of neighborhood area associated with the value occurring at the target point itself.

Another locational attribute which might be used, in conjunction with size, to characterize a neighborhood is cartographic distance from the target point. While distance has already been described as the basis for defining a neighborhood's absolute limits, it might be also be used to define the relative weights of values within a neighborhood. Noise level, for example, might be measured according to the inverse square of the distance from surrounding sources. The azimuthal relationship between a neighborhood location and a target point may also be used to weight the value associated with that location. In conjunction with distance weighting, this gives rise to a variety of sampling and interpolation techniques. Azimuthal relationships may also

be used to define absolute neighborhood limits.

### CARTOGRAPHIC MODEL

In order to suggest some of the ways in which primitive map processing operations might be combined to perform more complex analyses, an illustrative cartographic model is outlined below and schematically represented in Figure 5. The model provides an analytical procedure for integrating spatial information in assessing timber supply (Berry and Sailor, 1982). This supply is traditionally characterized solely in terms of standing timber. However, the relative accessibility and availability of each forested parcel must be considered in establishing effective supply. Maps of terrain characteristics and the road network are used to generate a map of accessibility. Fundamental to this analysis is the use of a neighborhood operation for conversion of a map of elevation into a map of topographic slope. A distance measuring operation is then invoked to establish the relative proximity



**Figure 5.** Flowchart of Effective Timber Supply Model. The cartographic model considers access and availability of forested areas, as well as physical inventory, in characterizing timber supply. Several primitive operations are logically sequenced to form the model.



of forest parcels to roads considering areas with steep slopes as harvesting barriers which must be circumvented. A distance operation is also used to establish "haul zones" from a mill based on the travel time along the road network.

The availability submodel uses a neighborhood operation to generate a map of housing density based on the total number of residential and commercial structures within a radius of 1/8 mile. A reclassification operation is used to establish the size of each ownership parcel. An overlay operation combines these two intermediate maps with one indicating areas excluded from harvesting to produce a map identifying the relative availability of areas for sale of stumpage. Areas of low housing density which are part of large ownership tracts are considered most likely to be available. The final submodel combines through an overlay operation, the maps of accessibility and availability to characterize effective timber supply. For selected combinations, a neighborhood operation is used to uniquely identify contiguous forest stands for management purposes.

Several other natural resource related models have been developed at the Yale School of Forestry and Environmental Studies using this approach as embodied in the Map Analysis Package software. These include:

- \* assessing deer habitat quality as a function of weighted proximity to natural and anthropogenic factors;
- \* mapping outdoor recreation opportunity as determined by an area's remoteness, size and physical and social attributes;
- \* predicting storm runoff from small watersheds by spatially evaluating the standard Soil Conservation Service model;
- \* assessing the spatial ramifications of the comprehensive plan of a small town considering natural land use, preservation, growth and utility policies; and,
- \* characterizing spatial relationships among marine ecosystems factors to model fish population dynamics.

In addressing these divergent applications a common set of fundamental map analysis operations were used. The logical sequencing of these operations on different sets of mapped data form the cartographic models of these different applications.

#### CONCLUSION

The modeling approach described in this paper can be used to extend the utility of maps for a variety of applications. A broad range of fundamental map analysis operations can be identified and grouped according to generalized characteristics. This organization establishes

a framework for understanding of the analytic potential of computer-assisted map analysis.

#### ACKNOWLEDGEMENTS

The research described in this paper was supported internal Yale University developmental research funds.

#### REFERENCES

- Berry, J.K. 1981. "Cartographic Modeling: Procedures for Extending the Utility of Remotely Sensed Data, "PECORA VII Symp. on Remote Sensing: An Input to Geographic Information Systems in the 1980's, American Society of Photogrammetry, 105 N. Virginia Avenue, Fall Church, Virginia, pp. 414-424.
- Berry, J.K. and J.K. Sailor. 1982. "Spatially Characterizing Effective Timber Supply," Proc. of The National Conference on Energy resource Management, Amer. Planning Asso., Vienna, Virginia.
- Berry, J.K. and C.D. Tomlin. 1982. "Cartographic Modeling: Computer-Assisted Analysis of Effective Distance, Proc. of Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana.
- Berry, J.K. and C.D. Tomlin. 1980. "Cartographic Modeling in the Classroom: A Fundamental Approach to Computer-Assisted Map Analysis," Proc. of the American Congress on Survey and Mapping, 40th Annual Meeting, pp. 227-241.
- Calkins, H. and D.F. Marble. 1980. Computer Software for Spatial Data Handling, volumes 1-3, International Geographical Union, Commission on Geographic Data Sensing and Processing, Ottawa, Ontario.
- Chrisman, N. and T. Peucker. 1975. "Cartographic Data Structures," The American Cartographer, Volume 2, Number 1, pp. 56-69.
- McHarg, I.L. 1969. Design with Nature, Doubleday/Natural History Press, Doubleday & Company, New York, 1969.
- Tomlin, C.D., In preparation. "Digital Cartographic Modeling Techniques in Environmental Planning," doctoral dissertation, School of Forestry and Environmental Studies, Yale University, New Haven, CT.
- Tomlin, C.D. and J.K. Berry. 1979. "A Mathematical Structure for Cartographic Modeling in Environmental Analysis," Proc. of the American Congress on Surveying and Mapping, 39th Annual Meeting, pp. 269-283.
- Tomlin, C.D. and S.M. Tomlin. 1981. "An Overlay Mapping Language Symposium on Regional Landscape Planning, American Society of Landscape Architects, Washington, D.C.

Part 3

TECHNIQUES, PROCEDURES AND DATA BASES:

POSTER SESSIONS

**Page intentionally left blank**

**Page intentionally left blank**



## PREPROCESSING OF THEMATIC MAPPER SIMULATOR (TMS) IMAGE DATA

Fred J. Gunther  
Computer Sciences Corporation  
Silver Spring MD 20910 U.S.A.

Belden G. Bly  
Computer Sciences Corporation  
Silver Spring MD 20910 U.S.A.

AND

William J. Campbell  
Eastern Regional Remote Sensing Applications Center  
NASA - Goddard Space Flight Center  
Greenbelt MD 20771 U.S.A.

Remotely-sensed images, raw from the sensor, are not suitable for computer-assisted analytical procedures. To get acceptably accurate results, distortions in the data must be removed before using computer-assisted procedures (i.e., classification mapping). Thus, Thematic Mapper Simulator (TMS) images need to be preprocessed before they can be analyzed. The TMS is a multispectral scanner with spectral bands that correspond to the Thematic Mapper (TM) on Landsat-D. TMS images were collected by aircraft and analyzed to demonstrate the benefits of classification mapping at higher spectral and spacial resolutions. This paper discusses the preprocessing procedures used and their results; the discussion has general application to any aircraft-collected, remotely-sensed imagery.

Preprocessing of TMS data is designed to do the following:

- o Remove geometric distortions. These are similar to the distortions in a picture produced by a wide-angle lense.
- o Remove radiometric distortions. These have the same cause as the geometric distortions, but include effects of the atmosphere, the angle of illumination, and slope.
- o Remove scan-line defects. These involve line-parallel image features (line drops, defective line starts, etc.).
- o Precision geometric correction. This allows direct, point-by-point comparisons between the image and a map, a geobased information file, or a different image.

Images recorded by aircraft scanner systems exhibit geometric distortions caused by altitude, ground speed, and foreshortening. The last is the "bow-tie" effect, in which pixels closer to each end of the scan lines represent progressively larger areas than those in the center. The image is corrected by resampling the data on a grid controlled by angular distance from nadir, altitude, and aircraft speed. The new image exhibits reasonably correct geometric relationships between objects within the image and those on the ground.

Correction for radiometric distortions involves an analytic approach to the relationship between look-angle and recorded reflectance values. This approach was devised by Mr. J. Irons (NASA-Goddard, Code 923) and Dr. M. Labovitz (NASA-Goddard, Code 922). A three-step procedure fits orthogonal polynomials to sampled image data and then applies a correction factor. The corrected image has reflectance values adjusted to what they would be if each and every pixel were taken directly below the aircraft.

Image defects involving many scan lines are fixed by processing the entire image. The image may be filtered or undergo Fourier transformation manipulations.

Image defects restricted to one or a few scan lines are fixed individually. The image or image statistics are searched for deviations from the general pattern. Depending upon the type of deviation, defective lines are fixed by replacement, by averaging replacement, or by subtracting a "fudge factor."

Precision geometric registration permits a very close correspondence between an image and a standard (a map or another image). The image is resampled according to a grid that is adjusted to control points selected in both the image and the standard. This compensates for numerous sources of error (e.g., changes in pitch, roll, yaw, ground speed, altitude).

Preprocessing is designed to present for analysis an image cleaned from systematic and accidental blemishes. It removes some sources of error that affect classification results.

## ABSTRACT

### DETERMINATION OF REMOTE SENSOR CAPABILITY BY MEANS OF AN AUTOMATED GEOGRAPHIC INFORMATION SYSTEM

R. Pascucci and A. Smith  
Autometric, Inc.  
Falls Church, VA 22047  
U.S.A.

A research project was performed under contract to the U.S. Geological Survey with the objective of determining the relative performance of five kinds of remote sensor imagery in the detection of geologic structure. The imagery that was examined was produced by: 1) a commercially-available, real-aperture, side-looking radar (SLR) system; 2) a commercially-available, synthetic-aperture SLR system; 3) standard-product black-and-white and color IR imagery from the Landsat Multispectral Scanning Subsystem (MSS); 4) digitally enhanced black-and-white and color IR imagery from the MSS; and 5) color photography from an aerial mapping camera.

The research methodology employed the following operations: 1) interpretation of the five remote sensor data sets for evidence of geologic structure, conducted by experienced remote sensing geologists; 2) digitization of the results of the interpretation and entry of the digitized data into an automated geographic information system, commercially known as AUTOGIS; and 3) manipulation of the digital data by the AUTOGIS to synthesize the data sets, to produce statistical tables showing the frequency and length of the geological structural features detected by each sensor, and to measure those features that were detected in common by two or more sensors and those that were uniquely detected by only one sensor.

The results of the AUTOGIS syntheses and analyses showed that, in respect to the total amount of geologic information detected, the sensors ranked, in descending order: real-aperture SLR system (most effective), digitally enhanced Landsat, aerial photos, synthetic-aperture SLR system, and standard-product Landsat (least effective). Equally surprising were the results showing the information detected uniquely and in common by combinations of two sensors. These showed that the amount of information detected in common by two-sensor combinations averaged only 26 percent of the total information detected by the two sensors, whereas the information detected uniquely by the two sensors used in combination averaged 74 percent. This means that on the average, when two of the sensors were interpreted, a net information increase of 59 percent ( $37/37+26$ ) was realized over and above the information that was obtained when only one sensor was interpreted.

The results of this investigation were obtained using a 10,000 square-mile test area in Alaska that had topography ranging from flat in the North Slope to rolling in the foothills of the Brooks Range. If the results should be shown to be more widely applicable in areas of different topography and vegetative cover, they could form the basis for cost-effective planning for remote sensor data acquisition in energy resource exploration



and in nuclear power plant siting.

The results have been used by the investigators in the production of an exploration aid that incorporates the synthesized results of real-aperture SLR and standard-product Landsat to produce three overlays showing: 1) lineaments; 2) fold axes, faults, and anomalies; and 3) favorable exploration targets within the Naval Petroleum Reserve, Alaska.

APPLICABILITY, COST, AND ACCURACY COMPARISONS  
OF SEVERAL CHANGE-DETECTION DIGITAL PROCEDURES

Fred J. Gunther  
Computer Sciences Corporation  
Silver Spring MD 20910 U.S.A.

AND

William J. Campbell  
Eastern Regional Remote Sensing Applications Center  
NASA - Goddard Space Flight Center  
Greenbelt MD 20771 U.S.A.

It is common experience that things change. Areas of change (dynamic areas) are extremely important to a resource manager. For a resource manager to efficiently detect and keep track of changes over a large area, time-sequential observations are needed. Landsat multispectral scanner system (MSS) images provide both time-sequential and wide-area views from space suitable for detecting many types of changes in surface cover over large areas.

Many different procedures have been proposed and used to detect changes using remotely-sensed data. Without field verification, changes in reflectance values recorded by a sensor and pointed out by computer-assisted image processing must be termed "alleged" changes. Testing of different procedures on anniversary image segments covering an area with four major landcover types (agricultural land, urban land, broadleaf forest, and waterbodies) indicated that different procedures are not equally valid or accurate. The application of different procedures to the same set of precision-registered images for central Pennsylvania has shown that they produce different numbers, locations, and proportions of alleged changes.

Several published procedures were found to be unsuccessful or unsatisfactory in detecting change. These procedures failed statistical tests or reconnaissance-level accuracy-assessment studies. They are: 1) Ratio Image Classification Differencing; 2) Ratio Image Difference Threshold Mapping; 3) Eigenvector Image Classification Differencing; 4) Eigenvector Image

Difference Classification; 5) Principal Components Analysis (2-mode Residuals) Threshold Mapping; and 6) Merged-Data Classification Mapping.

Four procedures were found to produce reasonable maps of alleged changes. These four procedures passed reconnaissance-level accuracy-assessment studies. Difference Image Threshold Mapping, a difference-delimite technique, was found to be the least costly (in terms of computer time and analyst time) of the procedures tested. The Difference Image Classification Mapping procedure also used little computer and analyst time. The PCA Residuals Classification Mapping procedure required nearly twice as much computer time as either of the previous two procedures. Post-Classification Change Detection was found to be both extremely labor-intensive (and therefore expensive) and costly in computer time. Detailed accuracy assessment studies of these four change-detection procedures is being conducted; results will be presented at the conference.



National Aeronautics and  
Space Administration

Washington, D.C.  
20546

Official Business  
Penalty for Private Use, \$300

THIRD-CLASS BULK RATE

Postage and Fees Paid  
National Aeronautics and  
Space Administration  
NASA-451



**NASA**

POSTMASTER: If Undeliverable (Section 158  
Postal Manual) Do Not Return